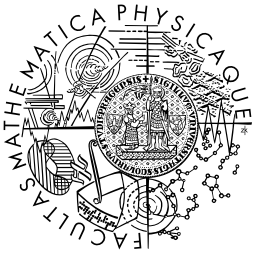


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Martin Veselý

Nonparametric tests of independence between animal movement trajectories

Department of Probability and Mathematical Statistics

Supervisor of the master thesis: RNDr. Jiří Dvořák, Ph.D.

Study programme: Mathematics

Study branch: Probability, Mathematical Statistics and Econometrics

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

I am thankful to my supervisor for his guidance and valuable advice. I also thank the Voyageurs Wolf Project for providing their data to be used in this thesis.

Title: Nonparametric tests of independence between animal movement trajectories

Author: Martin Veselý

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jiří Dvořák, Ph.D., Department of Probability and Mathematical Statistics

Abstract: In this thesis, we assume observing a pair of trajectories of two objects which could interact with one another and we want to propose a way to test their independence. We formulate basic point process definitions and discuss ways to describe trajectory data. We formulate the theory behind Monte Carlo tests and global envelope testing. In Chapter 2, we propose a parametric model to represent trajectories and derive Maximum Likelihood estimates of its model. We conclude the chapter by exploring the performance of these estimates. In Chapter 3, we propose test statistics used to test for independence using a nonparametric Monte Carlo test based on a random shift approach. We perform a simulation study to assess the performance of these statistics under various conditions and discuss the selection of fine-tuning parameters. Finally, in Chapter 4, we study real data provided by the Voyageurs Wolf Project and apply the proposed tests on real wolf trajectories.

Keywords: nonparametric statistics trajectory random shift Monte Carlo test

Contents

Introduction

This thesis is to a large degree motivated by data which was generously provided to us by the Voyageurs Wolf Project. We are interested in determining whether there is detectable dependence in the trajectories of tracked wolves in the Voyageurs National Park. To do this, we employ nonparametric tests of independence. Nonparametric tests were chosen due to their universality, because in the context of this problem there is a large amount of uncertainty about which model best describes the observed data.

In the first chapter, we provide basics of spatial statistics used in the rest of the thesis, including discussion about various ways to represent trajectory data, spatial process terminology, and theoretical results we use to devise the tests.

Then, in the second chapter, we define a model used to assess the performance of any proposed tests. To the best of our knowledge, no models have been proposed in literature covering animal tracking which are capable of modelling two interacting trajectories of animals, so we propose our own. Our model simulates data at discrete time points with equal times between observations, as this is also the nature of the data we have available to us. The model assumes a Markov property and needs to be able to generate independent trajectories, as well as various types and magnitudes of dependence between them. We focus especially on repelling interactions, since exploratory analysis of the real dataset suggests the interactions are more likely to be repelling than attracting. This is done by making the trajectories move independently in uniformly random directions until they are within a close enough distance of one another, at which point they prefer to move away from each other. Both the strength of this interaction and the barrier at which interaction beings are quantifiable parameters, which allows us to perform maximum likelihood estimation. The applicability of these results is also tested experimentally.

The testing is done using Monte Carlo testing using toroidal correction based on the works of Barnard [1963] and Lotwick and Silverman [1982]. Various test statistics need to be proposed and tested, both in terms of their achieved significance level and their power. Functional test statistics are explored as well, relying on the theory of global envelope tests from Myllymäki et al. [2017]. All simulations and computations are done using the software provided by R Core Team [2020].

In the fourth chapter we analyze the provided real data and finally apply the proposed tests on this data to determine whether we can find any significant dependence between the given wolf trajectories.

1. Basics of spatial statistics

In this chapter, we will define some basic terms and formulate the most important methods which we will use later on, especially in Chapter 3.

1.1 Ways to represent trajectory data

Throughout the entire thesis we will be dealing with data concerning trajectories of two objects (such as animals), and so a natural question is to consider how we can represent this data mathematically. For this section we will be following the notation and definitions in Schneider and Weil [2008].

1.1.1 Bivariate time series

Definition 1. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $T \subset \mathbb{R}$. The family of random variables $\{X_t, t \in T\}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$ is called a random (stochastic) process. If $T \subset \mathbb{N}_0 = \{0, 1, ...\}$, we call it a time series.*

Perhaps the simplest way to define a trajectory is to consider it a bivariate time series $((X_1, Y_1), ..., (X_n, Y_n))$, where X_i and Y_i are real-valued and represent the x and y coordinates of the object. This does, however, implicitly require the time intervals between the measurements to be equal, so we propose a more general way to represent trajectories.

1.1.2 Spatial-temporal point process

Let (E, ρ) be a separable metric space. The idea behind point processes is very simple. Imagine we were to randomly scatter a number of points on E and have $\Psi(B)$ be the number of points that landed in some set $B \subset E$. To arrive at the rigorous definition of such concept, however, we first have to define random measures.

Definition 2. *Let μ be a Borel measure on E and let $\mathcal{B}_0$ and $\mathcal{K}$ be the system of bounded Borel sets on E and the system of compact sets on E respectively. We say μ is locally finite if $\mu(K) < \infty \ \forall K \in \mathcal{K}$.*

We will then denote the space of all locally finite measures on $(E, \mathcal{B})$ by $\mathcal{M} = \mathcal{M}(E)$. We will also denote the space of all locally finite counting measures on $(E, \mathcal{B})$ by

$$\mathcal{N} = \mathcal{N}(E) = \{\mu \in \mathcal{M} : \mu(B) \in \mathbb{N} \cup \{0, \infty\} \ \forall B \in \mathcal{B}\}.$$

Definition 3. *For $B \in \mathcal{B}(E)$, let us define the mapping $\pi_B : \mathcal{M} \to [0, \infty]$ as $\pi_B(\mu) = \mu(B)$. We denote by $\mathfrak{M}$ the smallest σ-algebra on $\mathcal{M}(E)$ for which π_B is a measurable mapping $\forall B \in \mathcal{B}$, i.e.*

$$\mathfrak{M} = \sigma\{\pi_B \text{ measurable, } B \in \mathcal{B}\}.$$

Furthemore, we denote by $\mathfrak{N}$ the σ-algebra defined as the trace of $\mathfrak{M}$ on $\mathcal{N}$, i.e.

$$\mathfrak{N} = \{\mathcal{U} \cap \mathcal{N} : \mathcal{U} \in \mathfrak{M}\}.$$

With this knowledge we can finally move on to define a point process as a special case of a random measure.

Definition 4. *Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A random measure Ψ is a measurable mapping*

$$\Psi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{M}, \mathfrak{M}).$$

A point process Φ is a measurable mapping

$$\Phi : (\Omega, \mathcal{A}, \mathbb{P}) \to (\mathcal{N}, \mathfrak{N}).$$

A point process on $E = \mathbb{R}^d$ is called a *spatial point process*. A special case of a spatial point process on $E = \mathbb{R}^d \times T, T \subset [0, \infty)$, where T denotes time is called a *spatial-temporal point process*. If we choose $d = 2$, we get one of the possible ways we can formalize the concept of animal trajectories as a spatial-temporal process, where the observation (X, Y, T) represents an object at time T at coordinates (X, Y). The advantage of this approach is that the time difference between each pair of measurements $T_i, ..., T_n$ do not have to be equal. If the observed values are exactly $1, 2, ..., n$ the situation is equivalent to the definition proposed in Section 1.1.1.

1.1.3 Marked point process

Definition 5. *Let $\mathbb{M}$ be a complete separable locally compact metric space and call it the mark space. Denote its Borel σ-algebra by $\mathcal{B}(\mathbb{M})$. Let*

$$\mathcal{N}_m = \{\nu \in \mathcal{N}(\mathbb{R}^d \times \mathbb{M}) : \nu(\cdot \times \mathbb{M}) \in \mathcal{N}(\mathbb{R}^d)\}.$$

We say Φ_m is a marked point process if it is a point process on $\mathbb{R}^d \times \mathbb{M}$ such that

$$\mathbb{P}(\Phi_m \in \mathcal{N}_m) = 1.$$

The above definition gives us another way to represent the data we will be working with. We can consider trajectories to be a marked point process on $E = \mathbb{R}^2$ with marks from the mark space T which represent the time at which each observation was taken (we could also consider the mark space to be $\mathbb{R}^2$ and $E = T$, this seems less intuitive, however it provides the most natural context for the testing approach we take in Chapter 3).

1.2 Monte Carlo testing

The usual procedure for hypothesis testing is as follows: let us have $X_1, ..., X_n$ drawn from some distribution with parameters $\theta \in \mathbb{R}^m$ and a significance level $\alpha \in (0, 1)$ (a popular choice is $\alpha = 0.05$ but can be a lot lower than that depending on the specific problem). We want to test the hypothesis

$$H_0 : \ \theta = \theta_0,$$

against the alternative

$$H_1 : \ \theta \neq \theta_0.$$

We therefore take a *test statistic*

$$s_0 := S(\mathbf{X}),$$

where S is a measurable function and $\mathbf{X} = (X_1, ..., X_n)^T$ and derive its probability distribution under the assumption that H_0 holds. We would then take the probability of observing a value more extreme than the one we actually observed. For the sake of the example let us assume high values of S provide evidence against the null hypothesis (this corresponds to a one-sided test, a two-sided test can be constructed analogically, as can a one-sided test where low values of S provide evidence against the null hypothesis). If we denote by $\mathbb{P}_{\theta_0}(A)$ the probability of event A happening if the true value of θ is θ_0, this means we would compute

$$p = \mathbb{P}_{\theta_0}(S(\mathbf{X}) \geq s_0),$$

and reject H_0 if $p < \alpha$. The issue with this approach arises when we cannot analytically calculate the probability distribution of S. In those cases a so-called *Monte Carlo test* (Barnard [1963]) can be used, provided one can generate simulations from the model under the null hypothesis. We will still compute the test statistic $s_0 := S(\mathbf{X})$, but instead of finding its theoretical distribution, we will create N independent simulations, generating data from the model which satisfies H_0. From each of those simulations we will compute the same test statistic, so we end up with $s_1, ..., s_N$. We will then order them in ascending order to get $s_{(1)}, ..., s_{(N)}$ and compare them with s_0. Again, we will reject H_0 if s_0 is among the α most extreme values. If high values of S provide evidence against H_0, we will reject it if

$$s_0 > s_{(\lceil N(1-\alpha) \rceil)},$$

where $\lceil x \rceil$ is the so-called ceiling function defined as

$$\lceil x \rceil = \min\{n \in \mathbb{Z}, n \geq x\}.$$

Analogically, if small values of S are considered extreme, we reject the null hypothesis if

$$s_0 < s_{(\lfloor N\alpha \rfloor)},$$

where

$$\lfloor x \rfloor = \max\{n \in \mathbb{Z}, n \leq x\}.$$

If both large and small values provide evidence against H_0, we will reject it if

$$s_0 \notin \left(s_{(\lfloor N\frac{\alpha}{2} \rfloor)}, s_{(\lceil N(1-\frac{\alpha}{2}) \rceil)} \right).$$

It is obvious that a test constructed in such a way maintains the significance level α, because if the null hypothesis holds then all the random variables $s_0, s_1, ..., s_N$ are independent and identically distributed and so the probability of s_0 being in the most α extreme values is at most α (in fact it may be slightly lower depending on the exact value of N because of integer rounding). We can also guarantee the

test to have an exact significance level α if $\alpha(N+1)$ is an integer. Besag and Clifford [1989] also showed that the vector of test statistics $S_0, ..., S_N$ does not even have to be independent as long as they are exchangable, meaning

$$\mathbb{P}((S_0, ..., S_N) \in A) = \mathbb{P}((S_{\sigma(0)}, ..., S_{\sigma_N}) \in A)$$

for any measurable set $A \subset \mathbb{R}^{N+1}$ and any permutation σ.

For functional test statistics, the problem is more difficult. In the past, it has been suggested (for example by Ripley [1977] and Besag and Diggle [1977]) that if functional test statistics $S_0, ..., S_N$ are observed on an interval I, the empirical function $S_0(r)$ should be compared to the so-called *pointwise envelope*, i.e. the kth (most often for $k = 1$) smallest and largest values $S_i(r)$ for each $r \in I$. This of course is a classic example of the multiple testing problem so the significance level of this test will no longer be exactly α. In fact, it has been discussed (e.g. by Loosmore and Ford [2006]) that the significance level of such a test cannot be guaranteed given a predetermined choice of N and k. One way of working around the multiple testing problem is to consider so-called global envelope tests.

1.3 Global envelope tests

To extend the Monte Carlo test to functional statistics, we cannot rely on simply comparing the test statistics at individual points, we will have to define an ordering on the functions themselves. Let us denote such an ordering by $\prec$ (we will talk about a good choice of $\prec$ later). We will for the moment assume that $S_i \prec S_j$ means S_i is more extreme than S_j, of course the test will work analogically if the opposite holds instead. We will formalize the test using the formulation from Myllymäki et al. [2017, Lemma 3.1], while changing the notation to fit our needs.

Lemma 1. *Let $S_1, ..., S_{N+1}$ be exchangeable functional statistics, let $\prec$ be an ordering such that*

$$\mathbb{P}(S_i \prec S_j \vee S_j \prec S_i) = 1, \ \forall i \neq j,$$

and let

$$p = \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} \mathbf{1}_{(S_i \prec S_0)}\right). \tag{1.1}$$

Then, the test which rejects H_0 if

$$p \leq \alpha,$$

i.e.

$$1 + \sum_{i=1}^{N} \mathbf{1}_{S_i \prec S_0} \leq \alpha(N+1),$$

rejects the null hypothesis at the significance level α, provided that $\alpha(N+1)$ is an integer.

Proof. Let

$$A_j = 1 + \sum_{i=0, i\neq j}^{N} \mathbf{1}_{S_i \prec S_j}$$

be the rank of the functional statistic S_j among all the test statistics. The fact that no ties occur in the data with probability 1 implies that $(A_0, ..., A_N)^T$ is a permutation of $(0, ..., N)$ a.s. Then, for $j = 0, ..., N$, we have

$$\mathbb{P}(A_j = k) = \frac{1}{N+1}, \ k = 0, ..., N,$$

because the ranks $A_0, ..., A_N$ are exchangeable and hence uniformly distributed. Thus, it follows that

$$\mathbb{P}(p \leq \alpha) = \mathbb{P}(A_0 \leq \alpha(N+1)) = \frac{\alpha(s+1)}{s+1} = \alpha.$$

□

The equation 1.1 defines the p-value of the test, sometimes refered to as the Barnard p-value.

The question which arises is how to find a fitting ordering of functional data. In theory any functional ordering works, such as one based on functional depth (see e.g. Nagy, Stanislav et al. [2016]). We will focus mostly on *rank counts,* which is one of the most popular ways to order functional data in spatial statistics.

1.3.1 Extreme rank depth measure

We will now describe the ordering proposed by Myllymäki et al. [2017], which is called the *extreme rank depth measure*. Assuming for a moment that there are no ties in the values $S_i(r)$ for $r \in I$, we order the functions $S_0, ..., S_N$ according to the largest k for which they are still present in the k-th envelope, i.e.

$$R_i := \max\left(k : \ S_{\text{low}}^{(k)}(r) \leq S_i(r) \leq S_{\text{upp}}^{(k)}(r) \ \forall r \in I\right),$$

where $S_{\text{low}}^{(k)}(r)$ denotes the k-th lowest value out of $S_0(r), ..., S_N(r)$ and analogically $S_{\text{upp}}^{(k)}(r)$ denotes the k-th highest value.
Formally, we calculate the extreme rank depth measure as follows:

1. For each r, let $R_i^{\uparrow}(r)$ and $R_i^{\downarrow}$, $i = 0, ..., N$, denote the ranks of the values $S_i(r)$, $i = 0, ..., N$ from the smallest value to the largest, and from the largest to the smallest, respectively. In the case of ties in the values $S_i(r)$ for multiple i, we will use the mid-rank, although the maximum rank could also be used.

2. Let

   $$R_i^*(r) = \min\left(R_i^{\uparrow}(r), R_i^{\downarrow}(r)\right)$$

   denote the r-wise rank of $S_i(r)$. Smaller values of $R_i^*(r)$ signify that $S_i(r)$ is more extreme among the other functions at the point r. It is important to note this is the value we use when constructing a two-sided test. If we were trying to construct a one-sided alternative, we would only need to consider one of the values $\left(R_i^{\uparrow}(r), R_i^{\downarrow}(r)\right)$, rather than both of them.

3. We define the *extreme rank* as

$$R_i = \min_{r \in I} R_i^*(r).$$

We could now use R_i to order the functions S_i to perform the Monte Carlo test. The issue with this ordering is that the set of all R_i is guaranteed to contain ties because $\max(R_i) < N+1$, and therefore the functions $S_i(r)$ can be at most weakly ordered. Myllymäki et al. [2017] further suggests that a practical approach would be computing two different p-values,

$$p_- = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{1}_{(R_i < R_0)}, \tag{1.2}$$

$$p_+ = \frac{1}{N+1} \sum_{i=0}^{N} \mathbf{1}_{(R_i \leq R_0)}. \tag{1.3}$$

In this case p_- is a liberal p-value, and p_+ is a conservative p-value. This gives us a way to evaluate the test most of the time, since if $p_+ \leq \alpha$, we can clearly reject the null hypothesis, and if $p_- > \alpha$, we clearly do not have enough evidence to reject the null hypothesis. The issue arises when $p_- < \alpha < p_+$, since in that case it is not obvious at all whether to reject H_0 or not. The proposed solution is choosing p randomly in the p-interval. However, if we want to avoid this random choice, we can define an ordering which does not suffer from this problem.

1.3.2 Rank length and rank count

A better (in the sense that ties are a lot less likely to occur than with the extreme rank depth measure) ordering which has been suggested by Myllymäki et al. [2017] is the *rank length/rank count* ordering. Instead of simply taking the most extreme pointwise rank (R_i) that a function S_i attains in the interval I, we could still use the pointwise ranks but try to employ a more comprehensive summary of them. One such way would be to also consider how often the ranks $R_i^*(r) = k,\ k = 1, ..., \lfloor (N+2)/2 \rfloor$ have been attained by the function S_i. For a continuous function S_i, this can be expressed by a vector $\mathbf{L}_i$ of *rank lengths* L_{ik},

$$\mathbf{L}_i = (L_{i1}, ..., L_{i\lfloor (N+2)/2 \rfloor}), \quad L_{ik} = \int_I \mathbf{1}_{(R_i^*(r)=k)} \mathrm{d}r,$$

where L_{ik} symbolizes the "length" of the set $\widetilde{I} \subset I$ on which S_i was the k-th most extreme function. A natural way to order the functions S_i is then the so-called reverse lexicographic ordering

$$\mathbf{L}_i \prec \mathbf{L}_j \iff \exists n \leq \lfloor (N+2)/2 \rfloor :\ L_{ik} = L_{jk}\ \forall k < n,\ L_{in} > L_{jn}.$$

In other words, we order the vectors by their first element, in the case of any ties we try to break them using the second element, and so on. Of course, in practice we cannot calculate the integrals L_{ik} because we only observe the function at a discrete set of points. Let I_{fin} be the (obviously finite) set of points at which we evaluate the functions S_i. The vector of *rank counts* $\mathbf{C}$ is then defined as

$$\mathbf{C}_i = (C_{i1}, ..., C_{i\lfloor (N+2)/2 \rfloor}), \qquad C_{ik} = \sum_{r \in I_{\text{fin}}} \mathbf{1}_{(R_i^*(r)=k)}.$$

The ordering of rank counts is of course analogical to that of rank lengths. Obviously the ordering given by both the extreme rank lengths and the extreme rank counts is finer than that given by the most extreme ranks $R_i = \min_{r\in I} R_i^*(r)$ or $R_i = \min_{r\in I_{\text{fin}}} R_i^*(r)$, and it holds that

$$R_i < R_j \Longrightarrow \mathbf{L_i} \prec \mathbf{L_j} \wedge \mathbf{C_i} \prec \mathbf{C_j},$$

while the converse is not true (since if there was a tie in R_i, there were many other opportunities to break it given the ordering on $\mathbf{L}_i$ or $\mathbf{C}_i$).
Given the fact that the rank lengths L_{ik} are continuous, they define a strict ordering (i.e. one without ties) of the functions S_i as long as the condition $S_i \neq S_j,\ \forall i \neq j$ is met. The p-value p_L is obtained analogically to the equation 1.1 and the corresponding test which rejects H_0 if $p_L \leq \alpha$ has a significance level exactly α, according to Lemma 1, assuming again $\alpha(N+1)$ is an integer. The rank counts $\mathbf{C}_i$ are discrete, and thus ties may occur when ordering them. However, they are very unlikely for a reasonably sized set I_{fin}. If any ties do occur, however, they can be broken by randomization. That means, in practice, the Monte Carlo p-value p_N can also be obtained analogically to 1.1 as

$$p_C = \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} \mathbf{1}_{(\mathbf{C}_i \prec \mathbf{C}_0)}\right). \tag{1.4}$$

The following Lemma provides our own proof of Proposition 6.1 in Myllymäki et al. [2017] and summarizes the finding that a test based on rank counts is a refinement of the rest based on extreme rank depth.

Lemma 2. *Let p_- and p_+ be the p-values obtained by extreme rank ordering as defined in 1.2 and 1.3 and let p_C be the p-value obtained by rank count ordering as defined in* (1.4)*. Then*

$$p_- < p_C \leq p_+.$$

Proof. As we have mentioned before, it is obvious that

$$R_i < R_j \Longrightarrow \mathbf{C}_i \prec \mathbf{C_j},$$

and thus

$$\begin{aligned} p_- = \frac{1}{N+1}\sum_{i=0}^{N} \mathbf{1}_{(R_i<R_0)} &\leq \frac{1}{N+1}\sum_{i=1}^{N} \mathbf{1}_{(\mathbf{C}_i \prec \mathbf{C}_0)} \\ &< \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} \mathbf{1}_{(\mathbf{C}_i \prec \mathbf{C}_0)}\right) = p_C. \end{aligned}$$

Furthermore, it is also obvious from the way we defined the ordering $\prec$ that

$$\mathbf{C}_i \prec \mathbf{C_j} \Longrightarrow R_i \leq R_j,$$

which means that

$$\begin{aligned} p_C = \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} \mathbf{1}_{(\mathbf{C}_i \prec \mathbf{C}_0)}\right) &\leq \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} \mathbf{1}_{(R_i \leq R_0)}\right) \\ &= \frac{1}{N+1}\sum_{i=0}^{N} \mathbf{1}_{(R_i \leq R_0)} = p_+ \end{aligned}$$

□

1.4 Random shift approach

When dealing with Monte Carlo testing, we need to be able to generate values of the test statistic under the null hypothesis. However, the null hypothesis itself may not be enough to unambiguously define the entire model without further assumptions. For example, when observing the trajectories of two objects, we want the null hypothesis to be "The two trajectories are independent," rather than "The trajectories follow the exact model we described in Chapter 2 with $\kappa = 0$." However, we can only generate data under the latter assumption. To address this issue, Lotwick and Silverman [1982] suggests the following approach: Let us have two spatial processes, X and Y, and assume we want to test the hypothesis that these two processes are independent. Assuming the observation window is rectangular, we will identify the opposite edges and then wrap the data onto a torus. Then, keeping one of the processes fixed, we rotate the other process around the torus multiple times. By performing M such shifts, we receive M observations. The idea behind this approach is that by shifting one of the processes but leaving the other intact, any correlation structure which had existed between the two processes should not be present in the M generated observations, and thus the shifted processes $X_1, ..., X_M$ should be independent of Y, while any possible autocorrelation structure within X remains intact. Then we can compute the value of any test statistic for the shifted processes with Y remaining intact (i.e. $(X_1, Y), ...(X_M, Y)$), and use those values in the Monte Carlo test as the values of the test statistic under the null hypothesis.

The problem with this approach is that the described torroidal correction has the potential to break the autocorrelation structure inside of the shifted process, since the new process connects observations which do not belong together. For example, in our case, we plan on shifting the observations in time rather than in space, so such a shifted process could create two adjacent observations where the step taken to get from the first observation to the second one is larger than the model even allows. On top of this, such artifacts are larger for long shifts (i.e. shifting by a large number) than for short ones. This means the values $(S_0, S_1, ..., S_M)$ are not actually exchangeable after all, and the resulting test does not necessarily admit the correct significance level. Indeed, as was shown in Mrkvička et al. [2019], the toroidal correction often results in a very liberal test.

One way to deal with this issue is to apply the so-called minus correction. This corresponds to restricting the process X to a smaller observation window $W_C \subset W$, and then performing all shifts v in a way such that the shifted window remains fully within W, i.e. $W_C - v \subset W$. The test statistic is then calculated on $(Y|_{W_C}, X_i|_{W_C})$. The advantage is that this avoids the cracks in the autocorrelation structure caused by the torroidal correction. On the other hand, the issue is that we need a much larger sample size, since we are intentionally limiting ourselves to only using a smaller subset of the observations.

Another approach that also avoids breaking the autocorrelation structure, which has been proposed in Mrkvička et al. [2019], is the *variance correction*. This strategy relies on constructing a separate observation window for each shift

$v_1, ..., v_N$ as $W_i = W \cap (W + v_i)$. In other words, for each shift we simply remove the observations which would be placed outside the observation window. The test statistic S_i is then calculated on only the observations which remain, i.e. on $(Y|_{W_i}, X_i|_{W_i})$. Since each observation window W_i contains a different number of observations, the values $S_0, S_1, ..., S_N$ are no longer directly comparable, since the variability in the windows with more observations is lower than in those with fewer observations. This is solved by standardizing all the values of the test statistic to obtain values which have zero mean and unit variance:

$$\hat{S}_i = (S_i - \frac{1}{N+1} \sum_{i=0}^{N} S_i) / \sqrt{\text{var}(S_i)}, \qquad i = 0, ..., N.$$

The Monte Carlo test is then performed with the values $(\hat{S}_0, \hat{S}_1, ..., \hat{S}_N)$.

2. Trajectory model

In this chapter, we introduce the model which we created to simulate the trajectories of two interacting objects with the motivation of eventually testing the independence of two processes, which we will get to in the next chapter. We will approach the creation of this model in several steps, starting with the simplest variant and then generalizing it along the way.
To define even the simplest version of our model, we will first have to define the von Mises distribution, which is one of the most frequently used unimodal distributions for modelling angular data. We will be using the widely accepted definition from e.g. Mardia and Jupp [2000]:

Definition 6. *We say that a random variable $\theta \in (-\pi, \pi)$ follows the von Mises distribution with parameters $\mu \in (-\pi, \pi)$ and $\kappa \geq 0$, if*

$$f(\theta) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}, \qquad \theta \in (-\pi, \pi),$$

where

$$I_p(\kappa) = \sum_{r=0}^{\infty} \{(\Gamma(p + r + 1)\Gamma(r + 1)\}^{-1} (\frac{\kappa}{2})^{2r+p}$$

is the modified Bessel function of order 0. We will write $\theta \sim \mathrm{vM}(\mu, \kappa)$.

The von Mises distribution can be understood as an approximation of the normal distribution "wrapped" around the unit circle, where μ is the mean and $1/\kappa$ is analogous to the variance of a normal distribution, as can be seen in Figure 2.1. We can easily see that for $\kappa = 0$, the von Mises distribution is equivalent to a uniform distribution on the interval $(-\pi, \pi)$. For this reason some sources only allow the values $\kappa > 0$, but for our purposes it is easier to consider $\kappa = 0$ a valid value.

The general idea is to observe two separate bivariate time series $X_1, X_2, ..., X_n$, $Y_1, Y_2, ..., Y_n$ corresponding to two 2-dimensional trajectories which repel one another. First of all, let's describe the algorithm used to generate these trajectories, as that will naturally lead to an explanation of which parameters our model needs and what they mean:

1. We start by initializing X_0 and Y_0. For the purposes of the model it doesn't matter too much how we decide to initialize the starting points, they could be fixed or we could choose to generate the x and y coordinates independently at random to get $X_{0,x}, X_{0,y}, Y_{0,x}, Y_{0,y}$. Optionally, we also initialize X_1, Y_1 by shifting the points by a small constant in a random direction. This is only done for the purpose of allowing us to calculate angles with respect to the direction of the previous step, and could be avoided if we were to choose to calculate all angles with respect to the x-axis instead.

2. For $i = 1, 2, ..., n$, we do the following:

   (a) Generate $D_{X,i}, D_{Y,i}$ independently (both independent of each other and independent of previous observations) from a uniform distribution on $(0, R)$. This will serve as the distance by which we move each object.

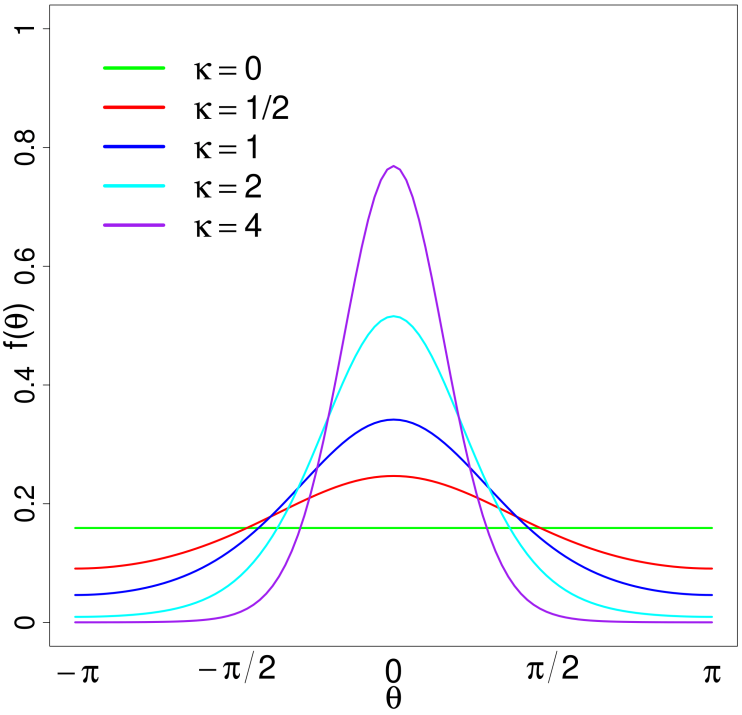


Figure 2.1: Density of the von Mises distribution for different values of κ

(b) Find the angle $\mu_{X,i}$, which would move the point X_i away from the point Y_i, as can be seen in Figure 2.2. Rigorously, we first take the vector $Z_i = Y_i - X_i$ and find the angle between Z_i and the vector $X_i - X_{i-1}$. This can be calculated as[1] $\mu_{X,i} = \text{atan2}(Z_{i,y}, Z_{i,x}) - \text{atan2}(X_{i,y} - X_{i-1,y}, X_{i,x} - X_{i-1,x})$. We then set $\mu_{Y,i} = \mu_{X,i} + \pi$. Note that if we had chosen to calculate the angles with respect to the x-axis instead, we would have obviously arrived at a different value but the resulting location of the new point would be the same.

(c) Generate $\theta_{X,i} \sim \text{vM}(\mu_{X,i}, \kappa)$ and $\theta_{Y,i} \sim \text{vM}(\mu_{Y,i}, \kappa)$

(d) Cenerate new points as follows:

$$X_{i+1} = X_i + \begin{pmatrix} \cos(\theta_{X,i}) & -\sin(\theta_{X,i}) \\ \sin(\theta_{X,i}) & \cos(\theta_{X,i}) \end{pmatrix} \cdot \frac{X_i - X_{i-1}}{||X_i - X_{i-1}||} \cdot D_{X,i},$$
$$Y_{i+1} = Y_i + \begin{pmatrix} \cos(\theta_{Y,i}) & -\sin(\theta_{Y,i}) \\ \sin(\theta_{Y,i}) & \cos(\theta_{Y,i}) \end{pmatrix} \cdot \frac{Y_i - Y_{i-1}}{||Y_i - Y_{i-1}||} \cdot D_{Y,i}.$$

This simply means we find the correct rotation for the new movement, normalize it to be the length of $D_{X,i}$ from step 2.(a) and add it to the last observed point location.

This model can easily be adapted to model attracting rather than repelling behavior by simply taking $\widetilde{\mu}_{X,i} = \mu_{X,i} + \pi$ in step 2.(c), as we show later on a more complex version of the model.

[1] As a reminder, the function $\text{atan2}(x, y)$ returns the angle between the point (x, y) and the x-axis, confined to the interval $(-\pi, \pi)$

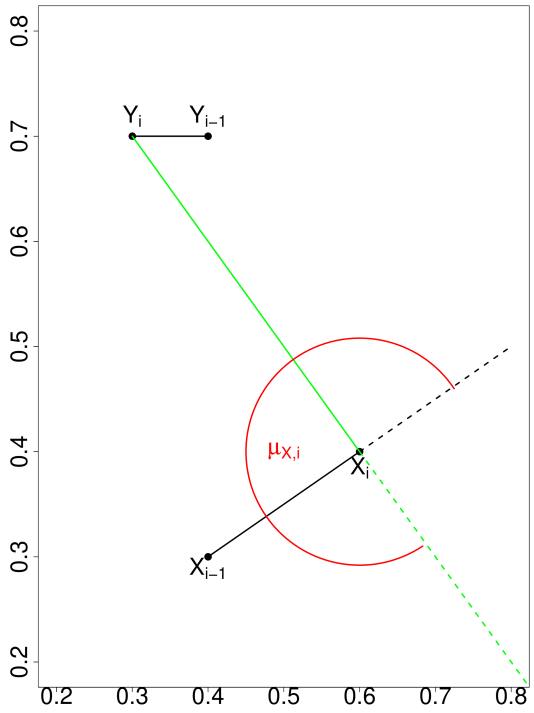


Figure 2.2: Illustration on how $\mu_{X,i}$ is chosen

From the construction of processes X_i, Y_i, we see a couple parameters that should be of note to us. The first one is R, which can be interpreted as the maximum possible length of each step, and κ as the parameter of the von Mises distribution used to generate θ, which can be interpreted as the strength of the interaction between X_i and Y_i. In theory, the two processes could have different parameters, but we will only consider the case where their distributions are the same. From the definition of the von Mises distribution, we get a lower bound for our first parameter $\kappa \geq 0$.

2.1 Model density and Maximum Likelihood

Beyond simply testing independence directly, another way of extracting information out of the observed data would be to estimate the parameters of the underlying model. For example, high estimates of κ can provide some level of evidence against independence of the two objects, as long as we know the estimate is precise enough. We will attempt to estimate the parameters in our model using maximum likelihood approach. Before we can do that, we will first need to find the density of the vectors $(X_1, ..., X_n)^T, (Y_1, ..., Y_n)^T$.

In the next section, we will for simplicity denote every density function simply by f, as long as it is obvious from the arguments which variable's density we are talking about.

First important thing to notice is that from the definition of our model, it is

obvious that the trajectory $(X_n, Y_n)^T$ has the Markov property in the sense that

$$\begin{aligned} &f(X_{k+1}, Y_{k+1}|X_k, Y_k; R, \kappa) = \\ &= f(X_{k+1}, Y_{k+1}|X_k, Y_k, X_{k-1}, Y_{k-1}, ..., X_0, Y_0; R, \kappa), \qquad \forall k \in 0, ..., n-1. \end{aligned}$$

It is also easy to see there is exactly one combination of $\theta_{X,k}, \theta_{Y,k}, D_{X,k}, D_{Y,k}$ that gets us from $(X_k, Y_k)^T$ to (X_{k+1}, Y_{k+1}). That means for each k, the points X_{k+1} and Y_{k+1} are unambiguously determined by the observations X_k, Y_k and the random variables $\theta_{X,k}, \theta_{Y,k}, D_{X,k}, D_{Y,k}$, so we can write

$$\begin{aligned} f(X_{k+1}|X_k, Y_k; R, \kappa) &= f(\theta_{X,k}; \mu_{X,k}, \kappa) \cdot f(D_{X,k}; R), \\ f(Y_{k+1}|X_k, Y_k; R, \kappa) &= f(\theta_{Y,k}; \mu_{Y,k}, \kappa) \cdot f(D_{Y,k}; R), \end{aligned} \tag{2.1}$$

and, setting $\mathbf{X_{k+1}} = (X_0, X_1, ..., X_{k+1})^T$, $\mathbf{Y_{k+1}} = (Y_0, Y_1, ..., Y_{k+1})^T$, and using the fact that

$$f(X_{k+1}, Y_{k+1}|X_k, Y_k) = f(X_{k+1}|X_k, Y_k) \cdot f(Y_{k+1}|X_k, Y_k),$$

we get

$$\begin{aligned} f(\mathbf{X_{k+1}}, \mathbf{Y_{k+1}}; R, \kappa) =& f(X_{k+1}|X_k, Y_k) \cdot f(Y_{k+1}|X_k, Y_k) \cdot f(X_k|X_{k-1}, Y_{k-1}) \\ &\cdot f(Y_k|X_{k-1}, Y_{k-1}) \cdot ... \cdot f(X_1|X_0, Y_0) \cdot f(Y_1|X_0, Y_0) \\ &\cdot f(X_0) \cdot f(Y_0). \end{aligned}$$

Rewriting the terms using (2.1) and reordering them, we get

$$\begin{aligned} f(\mathbf{X_{k+1}}, \mathbf{Y_{k+1}}; R, \kappa) = &\prod_{i=1}^{k} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \prod_{i=1}^{k} f(D_{X,i}; R) \cdot f(X_0) \\ &\cdot \prod_{i=1}^{k} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \prod_{i=1}^{k} f(D_{Y,i}; R) \cdot f(Y_0). \end{aligned} \tag{2.2}$$

While X_0, Y_0 are technically random variables, we don't want to specify their distribution, considering it could be basically arbitrary and it doesn't affect the other estimates. Thus we will treat this problem as finding conditional maximum likelihood conditioned on a realization of X_0, Y_0 (see for example Hamilton [1994]). Thanks to the product form of the density, we can estimate the parameters R and κ separately. The estimation of R is of very little interest to us, since the estimation of the parameter of a uniform distribution is rather trivial, and so according to e.g. Anděl [2007]

$$\hat{R}_{MLE} = \max_{\{i=1,2,...,n\}} \left(\max(D_{X,i}, D_{Y,i})\right),$$

i.e. the maximum of all observed step sizes from both trajectories.

The estimation of parameter κ is more interesting. The first issue we face is that technically each observation $\theta_{X,i}, \theta_{Y,i}$ is taken from a different distribution, since generally $\mu_{X,i} \neq \mu_{X,j}, i \neq j$. This also means the random variables $\theta_{X,i}, \theta_{Y,i}$ are dependent on all the previous observations of $\theta_{X,j}, \theta_{Y,j}, D_{X,j}, D_{Y,j}, j < i$, seeing as a different choice of distance changes the position of the two objects relative

to each other (and thus alters $\mu_{X,i}, \mu_{Y,i}$). Luckily for us, there's an easy way around this problem. We can notice that the density distribution of the von Mises distribution depends only on the term $\theta - \mu$, but not actually on the value of θ or μ individually. Thanks to the way we defined our model, $\mu_{X,i}$ is also a known parameter for each i (see step 2.(b)). This means we can simply "adjust" our observations to e.g. $\widetilde{\theta}_{X,i} = \theta_{X,i} - \mu_{X,i}$, and now $\widetilde{\theta}_{X,1}, ..., \widetilde{\theta}_{X,n}, \widetilde{\theta}_{Y,1}, ..., \widetilde{\theta}_{Y,n}$ represent $2n$ i.i.d. [2] random variables sampled from the distribution $\mathrm{vM}(0, \kappa)$, which are also independent of $D_{X,i}, D_{Y,i} \forall i$.
The problem then reduces simply to finding the maximum likelihood estimate of the parameter of a von Mises distribution. From this point onward, we will use the notation $N = 2n$, i.e. n is the length of each individual trajectory, whereas N is the total number of observations we have available.
Before we do that, we should first define some basic characteristic statistics used when working with angular quantities. Because a sample mean can be very misleading when sampling angles (see e.g. Jammalamadaka and Sengupta [2001] for examples), the so-called *circular mean* is often used instead.

Definition 7. *Let $\theta_1, ..., \theta_N$ be angular quantities. We define*

$$C = \sum_{i=1}^{N} \cos(\theta_i)$$
$$S = \sum_{i=1}^{N} \sin(\theta_i)$$
$$R = \sqrt{C^2 + S^2}.$$

The circular mean $\bar{\theta}_0$ *of observations $\theta_1, ...\theta_N$ is then defined as the solution to the equations*

$$\cos(\bar{\theta}_0) = \frac{C}{R}, \qquad \sin(\bar{\theta}_0) = \frac{S}{R}.$$

Remark. For the circular mean as defined above, it holds that

$$\bar{\theta}_0 = \arg\Big(\sum_{j=1}^{N} \cos(\theta_j) + i \sum_{j=1}^{N} \sin(\theta_j)\Big).$$

Theorem 3. *Let $\theta_1, \theta_2, ..., \theta_N \sim \mathrm{vM}(\mu, \kappa), \mu \in (-\pi, \pi), \kappa \geq 0$ be independent identically distributed random variables. Then the maximum likelihood estimates of μ, κ are given by*

$$\hat{\mu} = \bar{\theta}_0$$
$$\frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \sqrt{\Big(\frac{1}{N}C\Big)^2 + \Big(\frac{1}{N}S\Big)^2}, \tag{2.3}$$

where I_p is the modified Bessel function of order 0 as defined in Definition 6.

Proof. We know the density of θ_i is

$$f(\theta_i) = \frac{e^{\kappa \cos(\theta - \mu)}}{2\pi I_0(\kappa)}, \qquad i \in \{1, ..., n\}.$$

[2] independent identically distributed

Thus, the likelihood of $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)^T$ is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \frac{e^{\kappa \cos(\theta_i - \mu)}}{2\pi I_0(\kappa)},$$

and the log-likelihood is given by

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{N} \kappa \cos(\theta_i - \mu) - \sum_{i=1}^{N} I_0(\kappa) - 2\pi n.$$

Taking the partial derivatives with respect to μ and κ, and using the fact that $\frac{\partial I_0(x)}{\partial x} = I_1(x)$, which follows from differentiating the series expansion from Definition 1, we get

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \mu} = \kappa \sum_{i=1}^{N} \sin(\theta_i - \mu), \tag{2.4}$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \kappa} = \sum_{i=1}^{N} \cos(\theta_i - \mu) - N\frac{I_1(\kappa)}{I_0(\kappa)}. \tag{2.5}$$

Setting both equal to 0, we get

$$\kappa \cos(\mu) S = \kappa \sin(\mu) C,$$
$$N\frac{I_1(\kappa)}{I_0(\kappa)} = C\cos(\mu) + S\sin(\mu).$$

It is easy to verify these equations are solved by $\cos(\hat{\mu}) = \frac{C}{R}$ and $\sin(\hat{\mu}) = \frac{S}{R}$ (implying $\hat{\mu} = \bar{\theta}_0$ from the definition of circular mean) and $\hat{\kappa}$ satisfying (2.3).

□

Remark. The equation (2.3) does not have an analytical solution and must be approximated numerically (Mardia [1972]).

We can further improve this estimate using the knowledge we have about the parameter μ, so we are not forced to estimate both parameters at the same time, since for each $\widetilde{\theta}_i$ it holds that $\widetilde{\theta}_i \sim \mathrm{vM}(0, \kappa)$.

Theorem 4. *Let $\theta_1, \theta_2, ..., \theta_N \sim \mathrm{vM}(0, \kappa), \kappa \geq 0$ be independent identically distributed random variables. Then the maximum likelihood estimate of κ is given by the equation*

$$\frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{1}{N}\sum_{i=1}^{N} \cos(\theta_i). \tag{2.6}$$

Proof. As we have already shown in the previous proof, the partial derivative of the log-likelihood of the von Mises distribution is

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \kappa} = \sum_{i=1}^{N} \cos(\theta_i - \mu) - N\frac{I_1(\kappa)}{I_0(\kappa)}.$$

Considering $\mu = 0$ known, we immediately get

$$\frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = \frac{1}{N}\sum_{i=1}^{N} \cos(\theta_i),$$

which again can only be approximated numerically.

□

2.2 Generalizations of the model

The first generalization we will consider is adding a new parameter, for the moment assumed to be known, we will call it the *interaction distance* and denote it by I. Then we will only consider the interaction effect between X_i and Y_i if the distance between them is lower than I. Rigorously, this corresponds to

$$\theta_{X,i} \sim \begin{cases} \mathrm{vM}(\mu_{X,i}, \kappa), & \text{if } ||X_i - Y_i|| < I, \\ \mathrm{U}(0, 2\pi), & \text{otherwise,} \end{cases}$$

where $\mathrm{U}(0, 2\pi)$ is the uniform distribution on the interval $(0, 2\pi)$. For the density we now get

$$\begin{aligned} f(\mathbf{X_{n+1}}, \mathbf{Y_{n+1}}; R, \kappa; I) = &\prod_{i=1}^{n} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \prod_{i=1}^{n} f(D_{X,i}; R) \cdot f(X_0) \\ &\cdot \prod_{i=1}^{n} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \prod_{i=1}^{n} f(D_{Y,i}; R) \cdot f(Y_0) \\ = &\prod_{i:||X_i - Y_i|| < I} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \\ &\cdot \prod_{i:||X_i - Y_i|| < I} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \\ &\cdot \prod_{i=1}^{n} f(D_{X,i}; R) \cdot f(X_0) \cdot \prod_{i=1}^{n} f(D_{Y,i}; R) \cdot f(Y_0). \end{aligned}$$

Due to the density of a uniform distribution being constant, we know that

$$\begin{aligned} c_1 &= \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{X,i}; \mu_{X,i}, \kappa) = \left(\frac{1}{2\pi}\right)^{|i:||X_i - Y_i|| \geq I|} \\ &= \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \end{aligned}$$

(where by $|i : ||X_i - Y_i|| \geq I|$ we understand the number of indices i satisfying the condition $||X_i - Y_i|| \geq I$) is also constant with respect to the parameters $\mu_{\{X,Y\},i}, R_{\{X,Y\},i}, \forall i$. Thus we are simply trying to maximize

$$\begin{aligned} L(\kappa) = &\prod_{||X_i - Y_i|| < I} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{X,i}) \\ &\cdot \prod_{||X_i - Y_i|| < I} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{Y,i}) \end{aligned}$$

with respect to κ, which, after setting

$$\begin{aligned} \widetilde{\theta}_{X,i} &= \theta_{X,i} - \mu_{X,i}, \\ \widetilde{\theta}_{Y,i} &= \theta_{Y,i} - \mu_{Y,i}, \end{aligned}$$

yet again leads to the maximum likelihood estimation for the von Mises distribution. The only difference from the original model is that the sample size used to compute the estimate $\hat{\kappa}$ may be significantly smaller depending on the choice of I, so it will be less accurate if we don't compensate for this by increasing n.

The last generalization we will actually implement considers the same model we just described, but this time the parameter I is assumed to be unknown. Thus the density still remains

$$\begin{aligned}
f(\mathbf{X_{n+1}}, \mathbf{Y_{n+1}}; R, \kappa, I) = &\prod_{i=1}^{n} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \prod_{i=1}^{n} f(D_{X,i}; R) \cdot f(X_0) \\
&\cdot \prod_{i=1}^{n} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \prod_{i=1}^{n} f(D_{Y,i}; R) \cdot f(Y_0) \\
= &\prod_{i:||X_i - Y_i|| < I} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{X,i}; \mu_{X,i}, \kappa) \cdot \\
&\cdot \prod_{i=1}^{n} f(D_{X,i}; R) \cdot f(X_0) \cdot \prod_{i=1}^{n} f(D_{Y,i}; R) \cdot f(Y_0) \cdot \\
&\cdot \prod_{i:||X_i - Y_i|| < I} f(\theta_{Y,i}; \mu_{Y,i}, \kappa) \cdot \prod_{i:||X_i - Y_i|| \geq I} f(\theta_{Y,i}; \mu_{Y,i}, \kappa),
\end{aligned}$$

however we cannot easily compute the individual products since we don't know which observations fullfil the condition $||X_i - Y_i|| < I$. It is obviously not feasible to maximize this expression analytically, so we have to approximate it numerically. The most accurate way would be to order all N values $\delta_i = ||X_i - Y_i||$, then for each i take $I_i = \delta_i + \frac{\delta_i - \delta_{i-1}}{2}$, compute $\hat{\kappa}_i$ as the maximum likelihood estimate of the model which assumes $I = I_i$, compute the likelihood

$$L_i := f(\mathbf{X_n}, \mathbf{Y_n}; R, \hat{\kappa}_i, I_i) \tag{2.7}$$

for each i and then choose $\hat{I}_{MLE} = I_i$ for i which maximizes the likelihood function (2.7). This would however be very computationally demanding, especially for larger n, so we will have to be satisfied with only using a smaller grid of pre-selected I_i (we will elaborate on the selection in the simulation study chapter) and approximating $\hat{I}_{MLE}$ from this set.

The model can of course be generalized further. One option would be to consider interactions where the processes attract each other (either when they're close or when they're far away) - in this case the estimation would be exactly the same except all values of $\mu_{X,i}$ would be shifted by π. Those two models could also be combined, meaning the processes would repel each other when $||X_i - Y_i|| < I_1$ and attract each other when $||X_i - Y_i|| > I_2$ for some $I_1 \leq I_2$. The density of such model could be written as

$$\begin{aligned}
f(\mathbf{X_{n+1}}, \mathbf{Y_{n+1}}; R, \kappa, I_1, I_2) = &\prod_{i:||X_i - Y_i|| < I_1} f(\theta_{X,i}; \mu_{X,i}, \kappa_1) \cdot \prod_{i:I_1 \leq ||X_i - Y_i|| \leq I_2} f(\theta_{X,i}) \\
&\cdot \prod_{i:||X_i - Y_i|| > I_2} f(\theta_{X,i}; \mu_{X,i}, \kappa_2) \cdot \prod_{i=1}^{n} f(D_{X,i}; R) \cdot f(X_0) \\
&\prod_{i:||X_i - Y_i|| < I_1} f(\theta_{Y,i}; \mu_{Y,i}, \kappa_1) \cdot \prod_{i:I_1 \leq ||X_i - Y_i|| \leq I_2} f(\theta_{Y,i}) \\
&\cdot \prod_{i:||X_i - Y_i|| > I_2} f(\theta_{Y,i}; \mu_{Y,i}, \kappa_2) \cdot \prod_{i=1}^{n} f(D_{Y,i}; R) \cdot f(Y_0),
\end{aligned}$$

where we also allow for $\kappa_1 \neq \kappa_2$, meaning the strength of the two interactions can be different. The estimation of these parameters is still the same, with κ_1

being estimated only from the observations with a small distance, and κ_2 being estimated from the observations which are far apart.
The last generalization worth mentioning because of it's applicability is a variable interaction strength based on the distance between X_i and Y_i, meaning the particles attract each other more the closer they are to each other, so for distances $I_{j-1} \leq ||X_i - Y_i|| < I_j$, the interaction strength is κ_j. This can be written as

$$\begin{aligned} f(\mathbf{X_{n+1}}, \mathbf{Y_{n+1}}; R, \kappa, I_1, ..., I_m) = & \prod_{j=1,...,m} \prod_{i: I_{j-1} \leq ||X_i - Y_i|| < I_j} f(\theta_{X,i}; \mu_{X,i}, \kappa_j) \\ & \cdot \prod_{i: ||X_i - Y_i|| > I_m} f(\theta_{X,i}) \cdot \prod_{i=1}^{n} f(D_{X,i}; R) \cdot f(X_0) \\ & \cdot \prod_{j=1,...,m} \prod_{i: I_{j-1} \leq ||X_i - Y_i|| < I_j} f(\theta_{Y,i}; \mu_{Y,i}, \kappa_j) \\ & \cdot \prod_{i: ||X_i - Y_i|| > I_m} f(\theta_{Y,i}) \cdot \prod_{i=1}^{n} f(D_{Y,i}; R) \cdot f(Y_0), \end{aligned}$$

where for brewity we assume $I_0 := 0$ and the estimation follows the same logic as the previous models. The parameters $I_j,\ j = 1, ..., m$ could be known or estimated. An important consideration for the last two models is that the number of observations N required for accurate estimates is significantly higher when we have to estimate multiple different values of κ, considering we can only use a (potentially small) subset of data corresponding to each parameter.

So far the density and MLE we have derived are exact, however the model itself is very impractical. If we allow the objects to attain any arbitrary position in the plane $\mathbb{R}^2$, it is entirely possible the processes will simply drift apart for a very long time and we would need an extremely large number of observations to make any inferences about how the processes behave when they're close to each other. It is also not a realistic pattern of behavior for animals which tend to have specific territories, which were the motivation for our model. For this reason, it is more practical to restrict the processes to an observation window $(0, S) \times (0, S)$ for some $S > 0$ (other than square observation windows could of course be chosen as well). Note that it doesn't actually matter what S we choose because our other parameters, namely I and R, can be adjusted relative to the chosen S. To implement this, we will simply follow the steps 2.(a) - 2.(d) described above, but before moving to the next observation, we check to make sure that the new point lies in the observation window, i.e.

$$\begin{aligned} X_{i+1} &\in (0, S) \times (0, S), \\ Y_{i+1} &\in (0, S) \times (0, S). \end{aligned}$$

If not, there are several ways to correct the observation. The three most reasonable ones would be to allow the point to go outside of the observation window and bring it back in the next step, change D to not allow the point to leave the observation window, or change θ with the same goal. In our implementation of the model which we plan on using to test independence, we have decided to go with the third option. We simply set

$$\theta_{X,i,new} := \theta_{X,i} + \frac{\pi}{2},$$

or, in the case of Y_i,

$$\theta_{Y,i,new} := \theta_{Y,i} + \frac{\pi}{2},$$

and return to 2.(d). Note that after we generate the new point, we may have to rotate it again, up to a total of three times, which would correspond to a rotation by $\frac{3\pi}{2}$.
This adaptation creates another restriction for our parameters, namely we can only guarantee the algorithm will work if $R < \frac{S}{2}$ so that there is always a valid rotation that places the new point inside the observation window. This is only possible thanks to the fact that we are only considering rectangular observation windows, if a more complex choice of an observation window was made, we would need to be more careful. An unfortunate side-effect of this approach to making the model more applicable is that the random process X_i no longer has the density we just derived, however we will try to illustrate a way of at least partially correcting for this misspecification in the next section.

For the purposes of observing animal trajectories, it also makes sense to consider the fact that the observation window doesn't have to be the same for both objects, i.e. each object could have it's own territory in which it moves. This does not change the density or the maximum likelihood estimates in any way, but it's important to realize that yet again this has the potential to significantly reduce the number of observations where the objects are close together so a greater length of trajectories is required to make the same inferences.

2.3 Simulation study

In this part of the chapter, we implement the estimates devised in Sections 2.1 and 2.2 to see how accurate they are and how their quality changes for different values of parameters, mainly I, κ and the sample size N. First, though, we have to address the model misspecification caused by not including the out-of-bounds correction in the density. When deriving the maximum likelihood estimation, we assumed all N samples of $\widetilde{\theta}_{X,i}$ were generated from the distribution $\mathrm{vM}(0, \kappa)$. This is clearly wrong, considering some angles are likely going to be shifted by some multiple of $\frac{\pi}{2}$ to ensure they stay within the observation window. Considering the difficulty of including this effect in the density function, we will instead try to correct for it by removing all "suspicious" observations which could have potentially been a result of such correction. This means when computing the estimator we will not take into account observations for which

$$\mathrm{dist}(X_i, \partial((0, S) \times (0, S))) < D_{X,i},$$

where by ∂ we understand the boundary of a set, and "dist" is simply the distance between the point and the boundary. We will then compare this estimate with an estimate that has not been corrected (i.e. which has been computed with the incorrect assumption that every observation follows the desired distribution), and lastly with a so-called "oracle estimate", which has been obtained by keeping track of the originally generated $\theta_{X,i}$ (before an out-of-bounds correction has

been applied to it) during the data generation process and using those values to estimate κ to see whether the estimate remains reasonably accurate even after the imperfect correction, and also to show that the correction is needed because without it the estimators would be significantly more biased.
For the first set of experiments we performed simulations with the following parameters:

1. N = 1000, the combined length of trajectories (500 for each process),
2. S = 1, the length of the sides of the observation window,
3. R = 0.1, the maximum length of each step
4. M = 500, the number of simulations

and we compare the results for different values of κ and I. For now, I is assumed to be known. The parameter R is unknown, but as we stated previously, we do not consider the estimation of this parameter interesting so we will not estimate it. In Tables 2.1 and 2.2 we show the relative bias[3] of $\hat{\kappa}$, and relative mean square error of $\hat{\kappa}$ respectively, which are calculated as

$$\text{rBias}(\hat{\kappa}) = \frac{1}{M}\sum_{j=1}^{M}(\hat{\kappa}_j - \kappa)\cdot\frac{1}{\kappa},$$
$$\text{rMSE}(\hat{\kappa}) = \frac{1}{M}\sum_{i=1}^{M}(\hat{\kappa}_j - \kappa)^2\cdot\frac{1}{\kappa^2},$$

where $M = 1000$ is the number of independent repetitions of the experiment and $\hat{\kappa}_j$ is simply the maximum likelihood estimate of κ in the j-th simulation. In both tables we also include an information about the "true sample size" for the normal and oracle estimates, i.e.

$$\widetilde{N} = |i : ||X_i - Y_i|| \leq I \cdot S|.$$

It should be noted that the sample size used to calculate the corrected estimate is on average $\sim 10\%$ lower.

I	κ	non-corrected	corrected	oracle	true sample size
0.3	0.5	-0.0632	0.0454	0.0422	99
0.3	1.0	-0.0503	0.0253	0.0322	64
0.3	1.5	-0.0532	0.0354	0.0311	50
0.5	0.5	-0.1199	0.0293	0.0247	188
0.5	1.0	-0.0881	0.0210	0.0172	116
0.5	1.5	-0.0996	0.0101	0.0113	91
0.7	0.5	-0.2085	0.0032	0.0034	332
0.7	1.0	-0.1506	0.0062	0.0056	203
0.7	1.5	-0.1655	0.0033	0.0045	161

Table 2.1: Relative bias of $\hat{\kappa}$

[3]We have excluded 3 outlier estimates in total caused by a very small (i.e. ≤ 3) number of observations fulfilling $||X_i - Y_i|| < S \cdot I$ before calculating bias and MSE.

I	κ	non-corrected	corrected	oracle	true sample size
0.3	0.5	0.1205	0.1250	0.1159	99
0.3	1.0	0.0622	0.0628	0.0742	64
0.3	1.5	0.0543	0.0588	0.0518	50
0.5	0.5	0.0676	0.0600	0.0534	188
0.5	1.0	0.0369	0.0315	0.0297	116
0.5	1.5	0.0324	0.0226	0.0211	91
0.7	0.5	0.0665	0.0271	0.0251	332
0.7	1.0	0.0365	0.0161	0.0143	203
0.7	1.5	0.0392	0.0128	0.0119	161

Table 2.2: Relative MSE of $\hat{\kappa}$

We can see from the tables that while the MSE of the non-corrected estimate is not that bad (in fact, in some cases it was lower than the corrected and oracle estimates, due to the fact that it has less extreme outliers thanks to the negative bias and the estimates being bounded by 0), there is a very strong negative bias present for this estimate. This was to be expected, considering the out-of-bound correction is a lot more likely to change the observed angle in the exact opposite direction of μ than it is to shift the angle closer to μ, which makes the data seem a lot less concentrated around the mean (which would correspond to a lower value of κ).
Overall we can conclude the estimates are reasonable and work fairly reliably, at least for a large enough sample size. We did, however, also notice that for lower values of I and for higher values of κ, the number of observations which can be used for calculating the ML estimate is a lot lower than the length of the generated trajectories. This means if we were to select e.g. $N = 100$, which at first glance seems like a reasonable sample size, for $I = 0.3$ we would have practically no observations and couldn't compute $\hat{\kappa}$ at all.

For the second set of experiments, we will consider I an unknown parameter and try to estimate it together with κ. As mentioned previously, we cannot afford to compute the likelihood for every possible I so we will create a grid of 20 values and select the value of I which maximizes the likelihood (2.7). Ideally we would like the grid to be centered around the true value so we do not waste time computing L_i for values really far away from the true I, which would allow us to select a finer grid for values which are more likely to be selected by the method, however if we do this we risk artificially improving the estimate by artificially restricting it to "better" values. To avoid this, we will first create a coarse grid spanning the entire interval (0,1) and if the results are such that we can safely conclude the estimator doesn't miss the true value by a large margin, we will then perform a second set of simulations using a finer grid around the true value of I.

For the coarse grid, we have ran three experiments. In all three of them we have chosen $\kappa = 1$ and $N = 1000$, with the true value of I being $0.3, 0.5, 0.7$ in the three simulations. The admissible values for $\hat{I}$ are $0, 0.05, 0.1, ..., 0.95, 1.0$. We have only computed the corrected estimate in this simulation, skipping the normal and oracle estimates. As we can see from Figures 2.3, 2.4 and 2.5, the estimates

$\hat{I}$ are generally concentrated fairly well around the true value I. Denoting the percentage of estimates which have deviated from the true value of I by more than 0.2 by F, we have constructed the following table:

I	F
0.3	6.6%
0.5	3.1%
0.7	1.1%

These percentages are relatively small, therefore it seems perfectly reasonable to run the same estimates on a finer grid on the interval $(I-0.2, I+0.2)$ without artificially increasing the estimates' quality by removing a significant portion of incorrect estimates. An interesting fact to note is that the estimate $\hat{\kappa}$ is extremely reliant on choosing I accurately. For instance, when looking at the simulation for $I = 0.7$, the estimates $\hat{\kappa}$ corresponding to observations with $\hat{I} = 0.7$ had a rMSE of 0.013. The observations where $\hat{I} \neq 0.7$, on the other hand, produce estimates with a staggering $\text{rMSE}(\hat{\kappa}) = 1.193$ even after removing the most extreme outliers (i.e. $\hat{\kappa} > 10$). We illustrate the mean squared error of individual estimates in Figure 2.6.

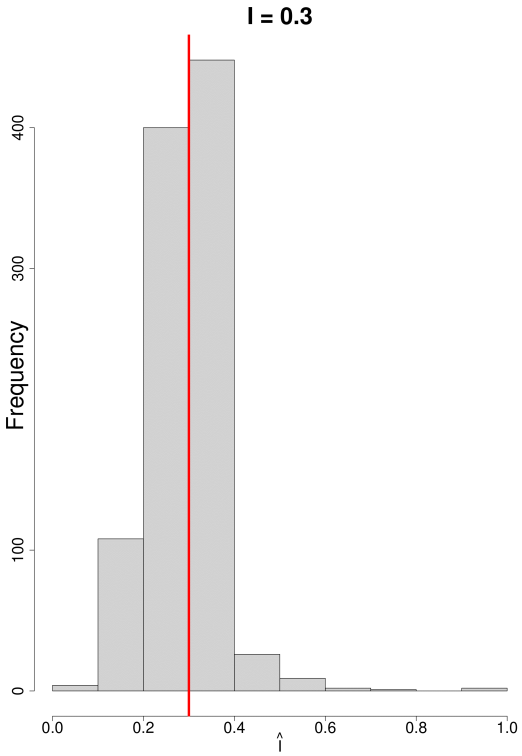


Figure 2.3: Estimates $\hat{I}$ for $I = 0.3$

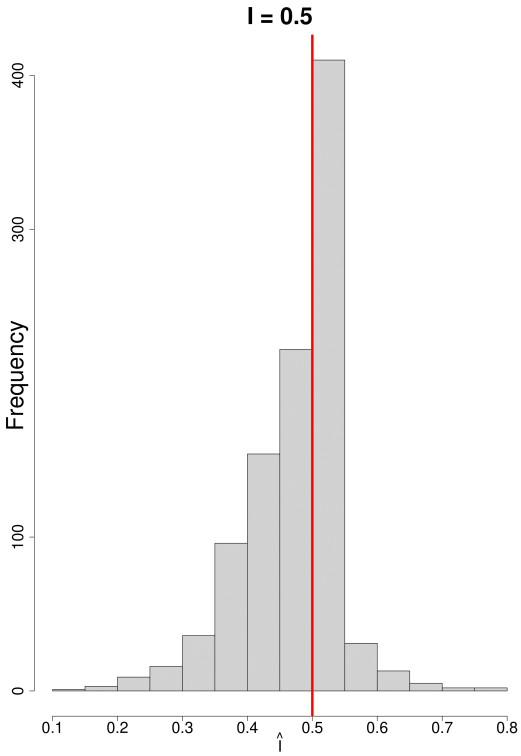


Figure 2.4: Estimates $\hat{I}$ for $I = 0.5$

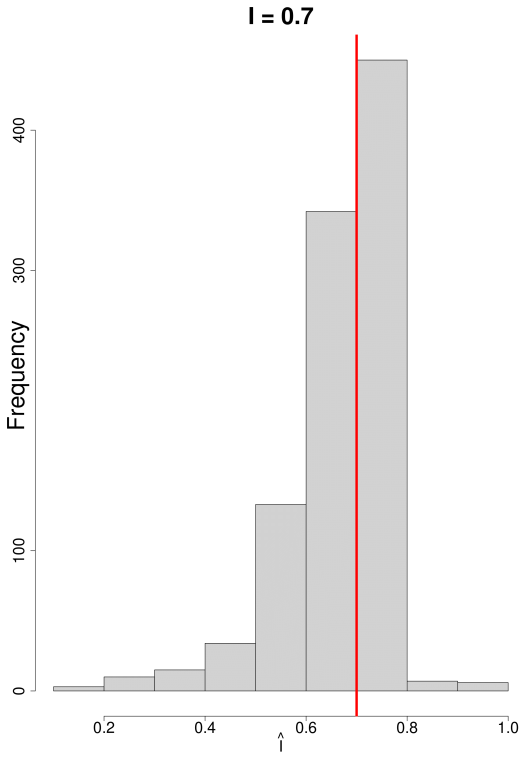
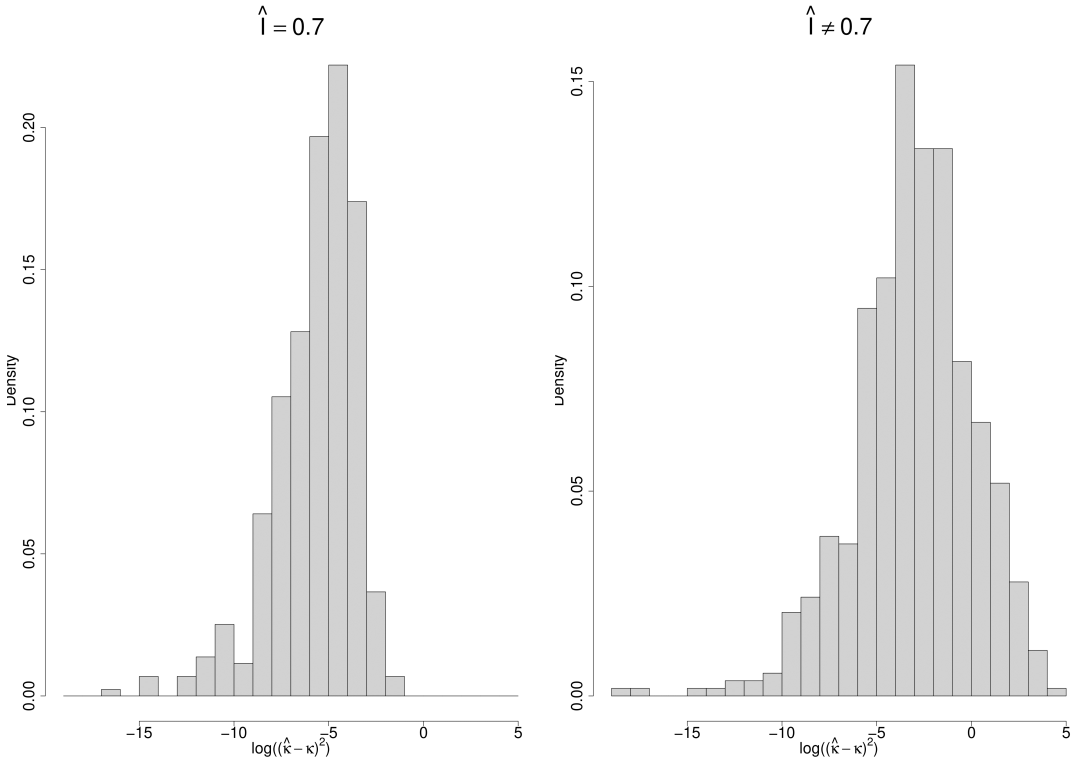


Figure 2.5: Estimates $\hat{I}$ for $I = 0.7$



(a) MSE($\hat{\kappa}$) for $\hat{I} = I$ (b) MSE($\hat{\kappa}$) for $\hat{I} \neq I$

Figure 2.6: Comparison of the MSE of the estimate $\hat{\kappa}$ based on the estimate of I (logarithmic scale was used due to varying magnitude)

For the simulations on a finer grid, we have again selected the parameters

1. n = 1000, the length of trajectories for each process (thus $N = 2000$ is the combined length)
2. S = 1, the length of the sides of the observation window,
3. R = 0.1, the maximum length of each step
4. M = 500, the number of simulations

and for each value I we will be computing the likelihood at points $\hat{I}_1 = I - 0.2, ..., \hat{I}_{21} = I + 0.2$. In Table 2.3 we show the relative bias and relative MSE of $\hat{\kappa}$, as well as the bias and MSE of $\hat{I}$. In Table 2.4 we display the percentage of estimates $\hat{I}$ which were "perfect", i.e. $\hat{I} = I$, and the percentage of estimates which were the worst possible estimates, i.e. $\hat{I} = I - 0.2$ or $\hat{I} = I + 0.2$.

I	κ	Bias(I)	rBias(κ)	MSE(I)	rMSE(κ)
0.3	0.5	-1.2674	1.5026	1.9827	9.3601
0.3	1.0	-0.3726	0.3807	0.3976	1.2439
0.3	1.5	-0.1420	0.1046	0.1359	0.3313
0.5	0.5	-1.4698	0.7465	2.4334	4.0559
0.5	1.0	-0.3720	0.2375	0.3991	0.7218
0.5	1.5	-0.1144	0.0046	0.0787	0.0886
0.7	0.5	-1.5922	0.2599	2.7167	0.8220
0.7	1.0	-0.3308	0.1056	0.4032	0.1808
0.7	1.5	-0.1052	-0.0344	0.0729	0.0363

Table 2.3: Bias and MSE of $\hat{I}$ and $\hat{\kappa}$

I	κ	perfect	worst
0.3	0.5	3.2%	15.7%
0.3	1.0	34.0%	1.0%
0.3	1.5	46.3%	0.0%
0.5	0.5	1.5%	26.4%
0.5	1.0	36.2%	1.9%
0.5	1.5	48.2%	0.0%
0.7	0.5	0.4%	29.1%
0.7	1.0	38.1%	1.3%
0.7	1.5	48.9%	0.1%

Table 2.4: Percentage of perfect and worst estimates $\hat{I}$

As we can see, the quality of the estimates of both I and κ decreases for lower κ. This was expected, considering the less the observations from the von Mises distribution differ from those sampled from the uniform distribution, the harder it becomes to distinguish between them and subsequently decide where the line separating them is. For $\kappa = 0.5$ the estimates seem practically unusable (at least without increasing sample size further). We can also see both the estimates improve slightly for higher values of I, which also makes sense since we saw the effect of I on the sample size in Table 2.1. It is also worth noting that the poor results for $\hat{\kappa}$ for $I = 0.3, \kappa = 1.0$ and $I = 0.5, \kappa = 1.0$ are mainly caused by a few outliers. After removing the observations where $\hat{\kappa} > 5$ (22 and 12 respectively, which corresponds to 2.2 % and 1.2 %), the rMSE would only be 0.5322 and 0.2806, respectively.

3. Trajectory independence testing

In this chapter, we will focus on the problem of testing the independence between two trajectories. One way to approach this would be to take the parametric model developed in Chapter 2, and then employing a parametric test with the hypotheses

$$H_0 : \kappa = 0$$
$$H_1 : \kappa > 0.$$

The issue with this approach lies in generalization. If we know exactly how the model behaves (because we were the ones who constructed it), we can estimate κ and try to formulate a parametric test. However, if we were to observe real life trajectories, we do not know the basics of the model that generated them. Even if we were to assume the data was generated exactly from the model described in Chapter 2, we would also need to have an idea of what the interaction barrier is, because as we saw in Chapter 2, the quality of $\hat{\kappa}$ decreases heavily when we do not have that information. Another issue is that we cannot simply apply known results about maximum likelihood estimates, since all the tests based on a maximum likelihood estimate (such as the Wald test or the likelihood ratio test) assume a regular density system. One of the characteristics of a regular density system is that the parametric space is an open set, which clearly is not the case in our problem since $\kappa = 0$ lies at the boundary of the parametric space.

Thus, it seems more practical to employ a non-parametric test, more specifically a Monte Carlo test using the random shift approach as described in Chapter 1. For this approach, we will adapt a different definition of a trajectory. Whereas in Chapter 2 we have considered it simply a bivariate time series (since we had $T = \{1, 2, ..., N\}$), the more general way would be to view it as a *marked point process*, where the marks represent the time at which the point was observed, as we discussed in Chapter 1.

Let us proceed to testing the independence of the trajectories of the model defined in Chapter 2. To clarify again, when talking about independence of trajectories, what we mean is that our null hypothesis is that X and Y are two independent processes which do not affect one another.

For our simulations, we simulate the "real" data (i.e. the processes which we want to test for independence) using the model defined in Chapter 2, and then we use the random shift approach with toroidal correction descibed in Chapter 1 to create replications of the processes which presumably satisfy the null hypothesis. This will, however, not be done by shifting the observations in space along the observation window W, but by shifting one of the processes in time. This is valid, because even though we consider the time of each observation to only be a mark of a process in $\mathbb{R}^2$, it would be equally accurate to view X, Y as processes in $T \times \mathbb{R}^2$ as discussed in Chapter 1 and so the random shift corresponds

to a special case of the random shift in $\mathbb{R}^3$ along one of the axes (or, even more naturally, we could consider the locations of the points to be marks and simply apply the shift to the real-valued process $T1, ..., T_n$). As we discussed in Chapter 1, the fact that we are utilizing the torus correction means our test is likely to be liberal to some degree. One way of dealing with this issue would be to apply the variance correction proposed by Mrkvička et al. [2019]. The complication with this approach is that computing the variance of the test statistic as a function of the size of the observation window is nontrivial. For this reason, we do not attempt to apply the variance correction for now, pending further investigation of the extent of the liberality introduced by the torus correction.

Considering the number of observed points in each simulation (i.e. the length of the trajectory of each process n) is fairly small (the longest trajectory we consider is 1000), it does not make much sense to perform an actual random shift, so we instead simply systematically iterate through all possible shifts to compute $S_1, ..., S_{n-1}$. Of course, the reason it even makes sense to talk about "all possible shifts" is that we are performing discrete shifts rather than continuous ones, as is the case in the general case of a random shift (because we only allow those shifts which result in the observation times of the shifted process and Y coinciding). This is analogous to performing permutation tests in classical statistics, where all permutations are systematically generated if the number of observations is not so large as to create the need for random permutations.

Again we will compare the results for different values of the parameters κ and the interaction barrier I. We will verify whether the test has the correct significance level, as well as investigate the power of the test based on multiple test statistics. We will compare both scalar and functional test statistics, using the simple Monte Carlo test for the scalar ones and the envelope tests for the functional ones. We will also look at the relationship between the estimate $\hat{\kappa}$ and the result of the non-parametric independence tests, with the expectation that the realizations where the null hypothesis is rejected will admit higher values of $\hat{\kappa}$.
We will consider the following parameters for all simulations:

- S = 1, the length of the sides of the observation window,
- R = 0.1, the maximum length of each step,
- M = 1000, the number of simulations.

All the test statistics we will be using will obviously be based on the distance between X and Y at a specific time. Let us denote this difference by

$$\Delta_i := ||X_i - Y_i||_2$$

and let the vector $(\Delta_{(1)}, ..., \Delta_{(n)})^T$ represent the values Δ_i sorted in ascending order. The three test statistics we will be using, denoted by "Mean", "Time", and "eCDF", are defined as follows:

- "Mean" - The mean distance Δ_i. However, since we are interested in the behavior of the processes when they are close to each other (because that

is when any interaction would appear under our assumed model), it would make sense to only compute the mean of the bottom $d \cdot 100$ % values. The effect of the choice of d will be explored to assess whether lower values perform better, and if so, until what point. This test statistic has the form

$$S_{\text{mean}} = \frac{1}{d \cdot n} \sum_{i=1}^{d \cdot n} \Delta_{(i)}.$$

- "Time" - The percentage of time when X and Y are close to each other, i.e. the number of observations D_i which are below a specified threshold c. This statistic seems like a good idea if we can easily select the threshold c, i.e. if we already have a prior assumption about what distance the two points have to be from each other before they start interacting with each other. Again, we will explore multiple choices of c. The test statistic has the form

$$S_{\text{Time}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(\Delta_i < c)}$$

- "eCDF" - The empirical distribution function of Δ. This is the only functional statistic out of the three described. It has the form

$$S_{\text{eCDF}}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(\Delta_i \leq t)}, \qquad t \in \left(0, \max_{i=1,\dots,n} (\Delta_i)\right).$$

  For the evaluation of this test we will use the global envelope test based on extreme length counts described in section 1.3.2.

3.1 Significance level with respect to different c and d

Let us first focus on the impact of our choice of d and c in the "Mean" and "Time" test statistic respectively. First we generate truly independent processes X, Y, which, given the model from chapter 2, is equivalent to generating processes with $\kappa = 0$. Then we display the observed significance level (i.e. the actual proportion of rejections) for $\alpha = 0.05$ for both test statistics and different choices of d and c to make sure the tests are not too liberal. We also show $\bar{\kappa}_0$ and $\bar{\kappa}_1$, which represent the average maximum likelihood estimates $\hat{\kappa}$ for those realizations where H_0 *was not* rejected, or where H_0 *was* rejected, respectively, to see how the results of the Monte Carlo test agree with the information about κ we can obtain from a maximum likelihood estimate. For this round of experiments we have considered trajectories with the length of $n = 500$.

We start with the test statistic S_{mean} :
As we can see from Table 3.1, the mean value of $\hat{\kappa}$ is indeed higher for the simulations where the null hypothesis was rejected, which gives us some confidence that the test is behaving as we expected it to. The mean of the estimates will, however, always be positive even though the true value of κ is equal to zero, since $\hat{\kappa}$ is bounded by zero, which leads to a positive bias as we have shown in Chapter

d	Observed significance level	$\bar{\kappa}_0$	$\bar{\kappa}_1$
0.1	0.080	0.019	0.030
0.2	0.072	0.019	0.029
0.5	0.081	0.019	0.032
0.7	0.079	0.019	0.035
1.0	0.091	0.019	0.034

Table 3.1: Observed significance level for different values of d in S_{mean}

c	Observed significance level	$\bar{\kappa}_0$	$\bar{\kappa}_1$
0.1	0.034	0.059	0.083
0.2	0.063	0.055	0.131
0.5	0.066	0.058	0.089
0.7	0.062	0.059	0.071
0.8	0.061	0.060	0.060
$\hat{I}$	0.052	0.057	0.111

Table 3.2: Observed significance level for different values of c in S_{time}

2. In terms of the observed significance level, we can see that all values of d result in a liberal test. As for assessing which value of the parameter results in the least liberal test, we cannot spot an obvious relationship (e.g. observed significance level increasing with increasing d), and while the observed significance level for $d = 0.2$ was the lowest, the difference between the various parameters does not appear to be statistically significant. For instance, if we assume the true significance level for all tests is 0.08 (which seems to be approximately the mean value), this corresponds to 80 rejections out of 1000 simulations with a standard deviation of approximately 9. That means both the lowest and highest observed significance levels are approximately within one standard deviation of the mean value.

Now we perform the same experiment for different values of c in the test statistic S_{time}. The one addition we have made compared to the previous tests is that we will also be trying to apply the test with $c = \hat{I}$, i.e. the maximum likelihood estimate of I, rather than a predetermined constant. After all, our hypothesis was that the test should perform better if c was equal to I, or at least reasonably close to it, so it appears reasonable to try and select this parameter based on the observed data. We need to be careful in this case though, because basing the test statistic on the data itself could introduce additional liberality, as the test may incorrectly interpret noise as signal more often. It is also important to note that the very idea of estimating I in an independent model is somewhat pointless, since for $\kappa = 0$ the value of I does not influence the likelihood function. However, we still need to assess the significance level of this option before applying it to models with dependent behavior (where the idea of estimating I is a lot more interesting).

From Table 3.2 we can see that the observed significance level is a lot lower for $c = 0.1$ and very similar across the other options. This is of course very suspicious, as the torroidal correction is always expected to make the test liberal,

or at the very least not this conservative. This finding could certainly be related to the fact that the values of this test statistic are very small (on average around 0.03, compared to approximately 0.12 for $c = 0.2$ and 0.5 for $c = 0.5$) so the variability is a lot greater. A positive finding seems to be that using $c = \hat{I}$ does not result in a more liberal test than setting c to a predetermined constant (at a first glance it actually appears *less* liberal). For illustration, we also show the estimates $\hat{I}$ in Figure 3.1. It should not come as a surprise that most of the estimates are small. This could also explain the smaller liberality of the test, with the lower estimates of $\hat{I}$ pushing the significance level a little bit lower.

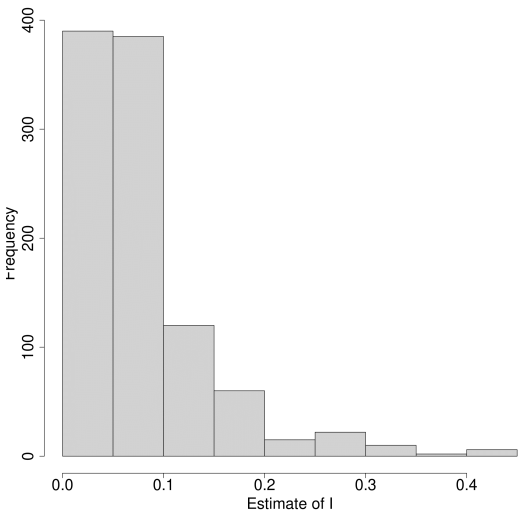


Figure 3.1: Histogram of $\hat{I}$ for independent processes

3.2 Power with respect to different c and d

We have explored how the significance level relies on the choice of c and d, so it is time to focus on how this choice impacts the power of the test. For the sake of brevity, we have chosen not to study all possible combinations of the values κ, I, n, d, but instead we split the experiments into two parts. First, we compare different choices of c, d when applied on a small number of models, and use this pilot experiment to select candidate values for c, d. We then proceed to a more detailed study where we compare the different test statistics with those pre-selected hyperparameters for a larger array of model parameters.

To assess the power of the test using the test statistic S_{mean} with respect to the parameter d, we will test observations generated by models with two different parameter combinations:

- $\kappa = 1.5, I = 0.5, n = 500,$
- $\kappa = 0.5, I = 0.7, n = 500.$

The idea is that in the second experiment we are significantly lowering the interaction strength but slightly increasing the interaction barrier to compensate for that.

d	Observed power	$\bar{\kappa}_0$	$\bar{\kappa}_1$
0.1	0.806	1.535	1.523
0.2	0.727	1.504	1.535
0.5	0.472	1.527	1.526
0.7	0.324	1.524	1.540
1.0	0.243	1.544	1.532

Table 3.3: Observed power for different values of d in S_{mean} with $\kappa = 1.5, I = 0.5$

d	Observed power	$\bar{\kappa}_0$	$\bar{\kappa}_1$
0.1	0.544	0.497	0.527
0.2	0.566	0.495	0.525
0.5	0.510	0.506	0.519
0.7	0.407	0.508	0.513
1.0	0.251	0.508	0.514

Table 3.4: Observed power for different values of d in S_{mean} with $\kappa = 0.5, I = 0.7$

From Tables 3.3 and 3.4, we can make several interesting observations. First of all, it seems that our hypothesis about the role of d was correct, and lower values of d (i.e. limiting ourselves to looking at the closest pairs of points) provide the test with a better ability to distinguish between dependent and independent pairs of processes. The power of the test is by far the highest for $d = 0.1$ and $d = 0.2$, and there is a clear drop-off in the power as d increases beyond 0.5.

We can also see the power was generally slightly higher for $\kappa = 1.5, I = 0.5$ than for $\kappa = 0.5, I = 0.7$, which is in contrast to Chapter 2, where the quality of the maximum likelihood estimates on the latter model was a lot higher than on the former one. This can easily be explained by the fact that the ML estimates are extremely reliant on the sample size, whereas in our approach to independence testing, we do not necessarily get more information by increasing the interaction barrier beyond a certain point (since we limit ourselves to only using the lowest $d \cdot 100\%$ observations anyway) and thus benefit from the increased interaction strength in the data more.

Finally, we should note that the mean estimate for κ does not seem to really be related to the result of our test, especially in the first experiment, as the estimates $\hat{\kappa}$ are still fairly high even in the simulations where the null hypothesis was not rejected. While disappointing, it is also not very surprising considering the test statistic measures a very different property to the estimate of κ (the distance between observations as opposed to directly measuring the angles between two steps).

Let us now turn our attention to the effect of c on the power of the test using the test statistic S_{time}. Again we have used the same parameter combination as before, i.e.

- $\kappa = 1.5, I = 0.5, n = 500$,
- $\kappa = 0.5, I = 0.7, n = 500$.

c	Observed power	$\bar{\kappa}_0$	$\bar{\kappa}_1$
0.1	0.004	1.538	1.823
0.2	0.061	1.530	1.552
0.5	0.665	1.522	1.525
0.7	0.014	1.530	1.464
0.8	0.021	1.538	1.494
$\hat{I}$	0.665	0.210	1.494

Table 3.5: Observed power for different values of c in S_{time} with $\kappa = 1.5, I = 0.5$

c	Observed power	$\bar{\kappa}_0$	$\bar{\kappa}_1$
0.1	0.001	0.518	0.596
0.2	0.022	0.513	0.494
0.5	0.493	1.426	0.517
0.7	0.248	0.504	0.530
0.8	0.018	0.516	0.472
$\hat{I}$	0.556	0.205	0.198

Table 3.6: Observed power for different values of c in S_{time} with $\kappa = 0.5, I = 0.7$

We have summarized the results in Tables 3.5 and 3.6.

As we can see in Table 3.5, for a large enough κ the test behaves exactly as we expected: For a c lower than the interaction threshold I, the test has practically zero power because the number of observations where $\Delta_i < c$ is extremely low even in the replicated processes. For $c > I$ however, we observe a large number of points with $I < \Delta_i < c$, so the power suffers yet again. On the other hand, it seems slightly surprising that for $\kappa = 0.5, I = 0.7$ (Table 3.6) the power is actually higher for $c = 0.5 \neq I$. We hypothesize this is because the interaction strength κ is small enough that the number of observations in which the distance between the two objects is less than I is fairly large, and limiting this to a more reasonable number provides higher power. An even better choice, however, seems to be the aforementioned $c = \hat{I}$. In the first experiment, it performed the same as $c = I = 0.5$, and in the second experiment it actually managed to outperform every predetermined constant we tried using.

Just like before, we see the value of κ is not related to the test's decision most of the time, and for $c = \hat{I}$ we even see values that are very far away from the actual value of κ. The latter observation can easily be explained by the fact that in Chapter 2 we demonstrated heavy reliance of the quality of $\hat{\kappa}$ on the correct specification of I.

Overall in terms of practicality, a test based on S_{mean} seems a lot more universally useful than a test based on S_{time}, since for the former we can simply choose d universally to be low (0.1 or 0.2 seem like good candidates), whereas the latter test is heavily reliant on c being chosen accurately to be reasonably close to I. That means, if we do not possess a priori information about I, the only plausible way of using the test based on S_{time} seems to be to use $c = \hat{I}$. This does, however, increase computational difficulty, and it has the potential to break down in the

n	Test statistic	Significance level	$\bar{\kappa}_0$	$\bar{\kappa}_1$
250	Mean	0.094	0.025	0.044
250	Time ($c = 0.5$)	0.082	0.025	0.043
250	Time($c = \hat{I}$)	0.051	0.026	0.035
250	eCDF	0.102	0.027	0.030
500	Mean	0.072	0.019	0.028
500	Time ($c = 0.5$)	0.064	0.019	0.038
500	Time($c = \hat{I}$)	0.059	0.019	0.029
500	eCDF	0.056	0.019	0.030
1000	Mean	0.050	0.013	0.031
1000	Time ($c = 0.5$)	0.059	0.013	0.024
1000	Time($c = \hat{I}$)	0.054	0.014	0.020
1000	eCDF	0.060	0.014	0.015

Table 3.7: True significance level for different test statistics

case of serious model misspecification.

3.3 Significance level with respect to sample size

From here on out we consider only the values $d = 0.2, c = \hat{I},$ and $c = I$ with the last one serving as some sort of a benchmark oracle model, which we do not have the chance to apply in real life situations (with the caveat that for an independent model there is no "correct" value of I, so for the significance level experiments we will use $I = 1$ to capture the fact that the movement pattern is homogenous at all distances) . For the first experiment, we want to demonstrate the observed significance level for various lengths of trajectories n. We saw both the tests were liberal for $n = 500$, so we would hope the liberality at least decreases with an increasing number of observations.

There are several interesting observations to make in Table 3.7. First of all, for most of the test statistics, the observed significance level gets closer to $\alpha = 0.05$ as the length of the trajectories increases, so the tests do get less liberal. The one exception is the test statistic S_{time} when using the parameter $c = \hat{I}$, which starts at an observed significance level of around 0.05 already for $n = 250$, and stays in the $(0.05, 0.06)$ interval as n increases. We can also see that the test based on an empirical distribution function is by far the most liberal for small sample size, but it does mostly catch up with the other test statistics as we increase n.
Overall, the liberality of the test (which, as we mentioned before, is caused by the toroidal correction) does not seem extreme enough to need to consider methods such as variance correction described in Mrkvička et al. [2019] as long as our sample size is reasonably large (e.g. $n \geq 500$).

I	κ	$S_{\text{mean, d=0.2}}$	$S_{\text{time},c=0.5}$	$S_{\text{time},c=\hat{I}}$	S_{eCDF}
0.3	0.5	0.235	0.154	0.182	0.140
0.3	1.0	0.254	0.198	0.225	0.138
0.3	1.5	0.256	0.207	0.219	0.168
0.5	0.5	0.316	0.190	0.252	0.175
0.5	1.0	0.423	0.309	0.345	0.176
0.5	1.5	0.461	0.388	0.402	0.204
0.7	0.5	0.340	0.127	0.301	0.191
0.7	1.0	0.524	0.322	0.422	0.248
0.7	1.5	0.523	0.400	0.425	0.264

Table 3.8: Power of the test, $n = 250$

I	κ	$S_{\text{mean, d=0.2}}$	$S_{\text{time},c=0.5}$	$S_{\text{time},c=\hat{I}}$	S_{eCDF}
0.3	0.5	0.302	0.239	0.312	0.114
0.3	1.0	0.403	0.402	0.457	0.117
0.3	1.5	0.414	0.464	0.461	0.145
0.5	0.5	0.544	0.338	0.530	0.206
0.5	1.0	0.713	0.588	0.654	0.246
0.5	1.5	0.741	0.699	0.721	0.228
0.7	0.5	0.581	0.244	0.517	0.309
0.7	1.0	0.789	0.573	0.702	0.364
0.7	1.5	0.852	0.720	0.729	0.341

Table 3.9: Power of the test, $n = 500$

3.4 Comprehensive power analysis

Now let us examine the power of these tests when applied to model with differing parameters. Similarly to Chapter 2, we consider all combinations of κ, I where $\kappa \in \{0.5, 1.0, 1.5\}$ and $I \in \{0.3, 0.5, 0.7\}$. We conduct one such experiment each for $n \in \{250, 500, 1000\}$, to also get an idea of how important the sample size is for different test statistics when it comes to power. We also stop providing $\bar{\kappa}_0$ and $\bar{\kappa}_1$, since the previous experiments showed there is not an interesting connection between this information and the rejection rate to justify making the table less readable.

As we can see, the power for $n = 250$ is fairly low, with even the most extreme parameter combination ($\kappa = 1.5, I = 0.7$) only achieving a power of at most 0.526. Generally the test statistic S_{mean} seems to perform the best for all possible combinations, with the other scalar statistic S_{time} achieving a $\sim 10 - 20\%$ lower power across the board. Slightly disappointing is the functional test statistic based on the empirical distribution function, which for the least extreme parameters achieves a power about the same as its significance level, and even for the higher values of κ, I achieves a power which is less than half of that of the scalar test statistics.

As we increase n to 500, the power of all the tests increases as expected. One change we can see is that at this point is that S_{mean} is no longer strictly superior to

I	κ	$S_{\text{mean, d=0.2}}$	$S_{\text{time},c=I}$	$S_{\text{time},c=\hat{I}}$	S_{eCDF}
0.3	0.5	0.527	0.461	0.697	0.152
0.3	1.0	0.691	0.743	0.815	0.181
0.3	1.5	0.723	0.818	0.835	0.207
0.5	0.5	0.862	0.621	0.873	0.362
0.5	1.0	0.931	0.862	0.907	0.368
0.5	1.5	0.943	0.931	0.932	0.372
0.7	0.5	0.871	0.494	0.864	0.483
0.7	1.0	0.958	0.840	0.901	0.554
0.7	1.5	0.988	0.942	0.937	0.522

Table 3.10: Power of the test, $n = 1000$

S_{time}, with the latter performing better for $I = 0.3$, and the former taking over yet again for higher values of the interaction barrier. Unfortunately, the functional test continues to be very underwhelming when it comes to power, even though we saw the significance level improve a lot at this sample size (which means at least the proportion of rejections in a dependent model versus an independent model is a lot better than it was for $n = 250$).

At $n = 1000$, the scalar test statistics put up respectable results even for lower values of parameters κ and I. Again we observe that S_{time} performs significantly better for low values of I, but S_{mean} achieves the highest power for stronger dependence between the two processes. While power over 0.9 seems very high, it is important to consider the extremeness of the simulated data at those parameters. We provide an example realization for $n = 1000, \kappa = 1.5, I = 0.7$ in Figure 3.2. We can see that in this case the interaction is so strong the processes almost seem like they are defined in two disjoint areas, and there are should be no doubts from looking at the data that the two processes are *not* moving independently in $(0, 1) \times (0, 1)$.

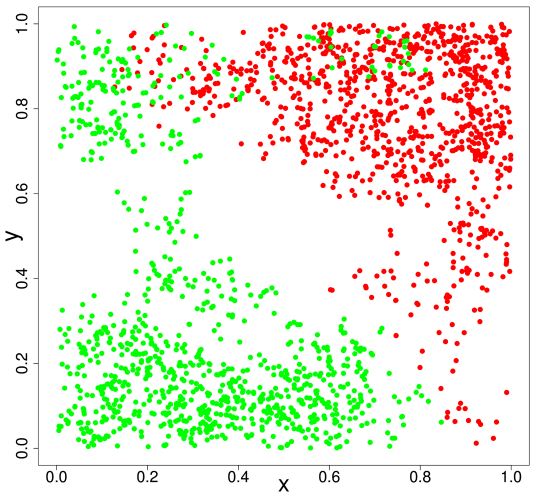


Figure 3.2: Realization of process X (green) and Y (red) for $n = 1000, \kappa = 1.5, I = 0.3$

n	eCDF	restricted eCDF
250	0.102	0.099
500	0.056	0.070
1000	0.060	0.045

Table 3.11: Significance level of the restricted eCDF test with $f = 0.2$ compared to the full eCDF test

Overall from these experiments it seems obvious that the scalar test statistics work a lot better than the empirical distribution function for all parameter combinations and across all sample sizes. The functional statistic for small trajectory lengths and/or small values of I, κ barely achieves power higher than its significance level, and even for the most extreme parameter combinations where we can achieve a power of ≥ 0.9 with the other 2 statistics, the evenlope test based on a cumulative distribution function attains at best half of that. This, compared to the fact that it suffers from by far the most extreme liberality for small sample sizes, makes us conclude there is no reason to employ this test statistic over the scalar ones. However, there is room to try and improve it, which we now explore.

3.5 Restriction of the empirical cumulative distribution function

So far we have seen the empirical cumulative distribution function perform very poorly as a test statistic for detecting dependence between two processes. A plausible hypothesis for why this is the case could be that the function looks at the *entire* dataset, while the scalar test statistics we have chosen ($S_{\text{time},c=\hat{I}}, S_{\text{mean},d=0.2}$) are able to look exclusively at the "interesting" subset of the data, i.e. the observations where the two processes are relatively close together. This certainly plays a nontrivial part in the power of the test, as we saw in Tables 3.3 and 3.4 where the observed power for $S_{\text{mean},d=0.2}$ was significantly higher than for $S_{\text{mean},d=1}$. Thus it seems to follow that a potential way to increase the power of the test based on an empirical cumulative distribution function is to only consider a restriction of the function:

$$S_{\text{eCDF}}(t) = \frac{1}{n}\mathbf{1}_{(\Delta_i<t)}, \qquad t \in \left(0, \Delta_{(f\cdot n)}\right),$$

for some constant f. Since all of our experiments in the previous section were performed with $d = 0.2$ when considering the test statistic S_{mean}, it seems fair to consider $f = 0.2$ as well, i.e. restricting the function to only the 20% lowest distances (We also tried $f = 0.5$ but the results were not interesting enough to report).

In Table 3.11 we can see the significance level does not appear to be significantly different from the original unrestricted eCDF significance level provided in Table 3.7. For $n = 500$ the test appears more liberal than before, but for $n = 1000$ it matches the nominal significance level well.

The power of the test is shown in Table 3.12. Immediately we can see the power is a lot higher than when we considered an unrestricted eCDF in Tables

I	κ	$n = 250$	$n = 500$	$n = 1000$
0.3	0.5	0.129	0.209	0.423
0.3	1.0	0.150	0.246	0.613
0.3	1.5	0.159	0.279	0.656
0.5	0.5	0.201	0.317	0.660
0.5	1.0	0.200	0.423	0.822
0.5	1.5	0.228	0.455	0.823
0.7	0.5	0.224	0.384	0.696
0.7	1.0	0.337	0.577	0.874
0.7	1.5	0.321	0.594	0.908

Table 3.12: Power of the restricted eCDF test , $f = 0.2$

I	κ	$S_{\text{mean, d=0.2}}$	$S_{\text{time},c=\hat{I}}$
0.1	1.0	0.091	0.056
0.1	2.0	0.079	0.062
0.2	1.0	0.177	0.242
0.2	2.0	0.231	0.322

Table 3.13: Power of the , $n = 500$

3.8 to 3.10. In fact, the power has at least doubled in all the parameter combinations, and the increase in power has been even greater in cases where I was low. However, even this is not enough to fully match the power of the scalar test statistics. The biggest gap seems to remain at small sample sizes, since for $n = 250$ the S_{mean} and S_{time} statistics achieved power in the range of 0.22 to 0.52 and 0.18 to 0.42 respectively depending on the parameter combination, while the restricted eCDF function lags behind at 0.13 to 0.32.

3.6 Detecting subtle interactions

In terms of comparison between the scalar statistics S_{mean} and S_{time}, it seems that the latter should be chosen when we hypothesize the interaction distance is lower and thus the interaction is more subtle, which seems more useful in most practical applications than being able to maximize power for the most extreme dependence. Let us explore this hypothesis a little further. For now we will set $n = 500$ as the tests seem to work fairly well at this sample size, but the calculations are a lot more feasible to perform than for $n = 1000$. We will try to check whether the superiority of S_{time} over S_{mean} still holds for even lower values of I. We consider $I \in \{0.1, 0.2\}$ and $\kappa \in \{1.0, 2.0\}$, which is slightly higher to compensate for the very low interaction barrier.

We can see the power for $I = 0.1$ is practically the same as the significance level when it comes to S_{time} and only slightly higher for S_{mean}, so the test does not work well. However, for $I = 0.2$, we observe decent power, especially for S_{time} which performs significantly better than S_{mean}. For illustration, we provide a graphical representation of a realization that the test using S_{time} was able to reject ($p = 0.04$). In Figure 3.3 we see all the locations of the process, and Figure

3.4 shows a histogram of distances between the two processes measured at each time instance. Even though the histogram shows some signs of interaction (e.g. no values $\Delta_i < 0.1$), we would probably have more trouble reaching a conclusion as confidently as before in Figure 3.2 without conducting a statistical test.

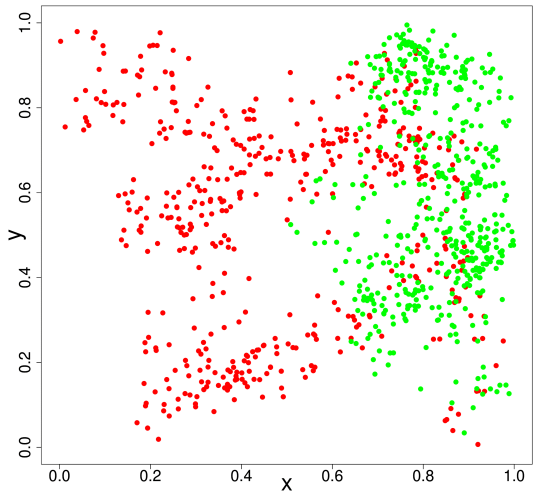


Figure 3.3: Realization of process X (green) and Y (red) for $n = 500, \kappa = 2.0, I = 0.2$

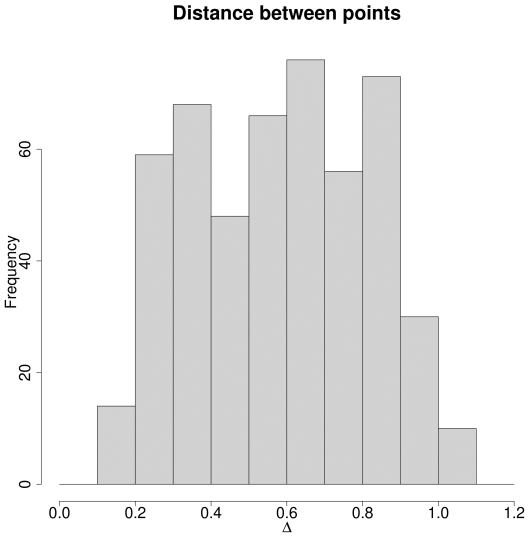


Figure 3.4: Histogram of distances between processes X, Y for $n = 500, \kappa = 2.0, I = 0.2$

3.7 Separate observation window for each process

As a final experiment, we consider two processes which occupy two disjoint observation windows W_1 and W_2, as this is the model which best describes the real world data we will look at in the following chapter. It seems clear that since the

n	$S_{\text{mean, d=0.2}}$	$S_{\text{time},c=\hat{I}}$	$S_{\text{eCDF},e=0.2}$
250	0.078	0.066	0.100
500	0.048	0.034	0.073
1000	0.062	0.054	0.059

Table 3.14: Observed significance level for processes in separate observation windows

processes can interact only when they are both close to the boundary between W_1 and W_2 (rather than being able to "meet" in any part of the observation window like before), we will need to increase either the sample size or the interaction barrier to achieve power comparable to the results from the previous experiments. For all the experiments we assume the following:

- $W_1 = (0,1) \times (0,1), W_2 = (1,2) \times (0,1)$ the two adjacent observation windows,
- M = 1000, the number of simulations.

We begin by exploring the observed significance level of the various statistics. The results are shown in Table 3.14. We can see that the results are mostly consistent with the experiments performed in a single observation window. The restricted empirical cumulative distirbution function yet again seems to be the most liberal test statistic. For $n = 500$ we see surprisingly low significance levels for the scalar test statistics, with the S_{time} statistic actually being conservative, but upon further investigation this seems to simply be caused by the variance in a particular batch of simulations (we did not see such conservative behavior when using a different random initialization).

To assess the power of the test, we will use similar parameter combinations as in Section 3.4:

- n = 500,
- $\kappa \in \{0.5, 1.0, 1.5\}$,
- $I \in \{0.5, 1.0, 1.5\}$.

Again, we have increased the values of I (from the original $\{0.3, 0.5, 0.7\}$) because of the larger combined area where the points are moving. Another option would be to significantly increase n, however this is computationally very difficult. We present the observed power in Table 3.15.

We can see that for $I = 0.5$ the test achieves a very small power (though still significantly higher than the significance level), which makes sense since the number of observations within that distance of each other is very low. This is not only caused by the fact that the maximum distance between points is greater than before, it is also caused by the fact that the "zone" in which the processes can even be in proximity to one another is very limited. Informally speaking, one of the processes has to be in the right part of its observation window at the same time as the other process is in the left part of its observation window. This is in contrast to the previous models, where the first process could be placed at an arbitrary point and there was a non-zero probability of the second process being reasonably close.

For higher interaction barriers, we observe higher achieved power by all test statistics. It should, however, not surprise us that this power is still smaller than it was for $n = 500$ in the previous simulations. The restricted empirical cumulative distribution function yet again achieves the smallest power in most cases, but actually performs better than S_{time} when I is at its highest and κ is low. It technically also achieves higher power for $I = 0.5, \kappa = 0.5$, but that power is barely higher than its significance level so its value in this case is questionable at best.

As far as the effect that we observed in Section 3.6 (i.e. the fact that S_{time} had higher power for smaller values of I), this effect was observed in smaller experiments for example for $I = 0.75$ (with the power being somewhere around 0.15 higher for S_{time} than for S_{mean}), but these results were not included in Table 3.15 to avoid cluttering the table with too many parameter combinations.

I	κ	$S_{\text{mean, d=0.2}}$	$S_{\text{time},c=\hat{I}}$	$S_{\text{eCDF},e=0.2}$
0.5	0.5	0.097	0.072	0.086
0.5	1.0	0.106	0.089	0.081
0.5	1.5	0.105	0.093	0.078
1.0	0.5	0.349	0.284	0.194
1.0	1.0	0.428	0.427	0.238
1.0	1.5	0.452	0.442	0.261
1.5	0.5	0.416	0.262	0.414
1.5	1.0	0.648	0.574	0.577
1.5	1.5	0.731	0.698	0.584

Table 3.15: Power of the test for two separate observation windows, $n = 500$

3.8 Simulation study summary

To summarize, the key takeaways from this chapter are:

- All of the tests are liberal to a certain degree due to the torroidal correction. The liberality decreases for increasing sample sizes, but one should be careful with small datasets.
- The cumulative distribution function did not perform well as a functional test statistic in any experiment.
- Restricting the empirical cumulative distribution function to a smaller interval around 0 makes it somewhat competitive with scalar statistics. It did not, however, perform better than them in any experiment. Due to this, and the added computational difficulty, we do not see a reason to use it over the two proposed scalar statistics.
- S_{time} can perform very well but is extremely reliant on an appropriate choice of c. This can be mostly worked around using a maximum likelihood estimate of I, but it remains uncertain how the test would perform on generic data that does not follow the model specification given in Chapter 2. A more sophisticated choice of c might be considered, such as limitting the permissible values of $\hat{I}$ such that a sufficient number of observations still lies below c.
- S_{mean} performs well as a universal choice of a test statistic that does not rely on any parametric representation of the underlying model, but it does not perform as well at detecting small scale interactions as S_{time}.

4. Application on the Voyageurs Wolf Project dataset

The data we are using in this chapter was generously provided to us by the Voyageurs Wolf Project. More information about the project along with interesting visualisations and animations can be found on the project website[1].
In this chapter we describe this data, and apply the tests developed in Chapter 3 to see if we can detect any type of statistically significant dependence between the trajectories of the tracked wolves from neghbouring packs.

4.1 The data

The monitored area contains six different wolf packs. In each of these packs, exactly one wolf was selected and equipped with a GPS tracker. The provided data thus consists of GPS coordinate measurements from these six trackers which documented their location every 20 minutes between April 15th 2018 and November 9th 2018. The files contain the following information:

- ID of the tracked wolf,
- Date and time of the tracker ping,
- Latitude of the tracker at the given time,
- Longitude of the tracker at the given time.

The set of all wolf packs provided to us and their respective IDs is summarized in the table below:

ID	Wolf pack
28	Moose River Pack
62	Bowman Bay Pack
64	Sheep Ranch Pack
66	Fawn Crick Pack
71	Lightfoot Pack
72	Moonshadow Pack

When refering to the individual wolf packs, we will simply refer to them by their ID, e.g. when we say "Wolf 64", we mean the GPS tracker assigned to a wolf from the Sheep Ranch Pack.

The first issue we have to deal with is that Latitude and Longitude do not form a Cartesian coordinate system on a plane, but rather they are spherical coordinates. That means finding the distance between two points is not as easy as finding the euclidean norm of the difference of their coordinates. The actual distance can be calculated using the so-called haversine function (see e.g. Sinnott

[1] https://www.voyageurswolfproject.org

[1984]):

Lemma 5. *Let (ϕ_1, λ_1) be the latitude and longitude of point A, and (ϕ_2, λ_2) the latitude and longitude of point B given in degrees. The great-circle distance between point A and point B in kilometers can be calculated as*

$$\rho(A, B) = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\hat{\phi}_2 - \hat{\phi}_1}{2}\right) + \cos(\hat{\phi}_1)\cos(\hat{\phi}_2)\sin^2\left(\frac{\hat{\lambda}_1 - \hat{\lambda}_2}{2}\right)}\right),$$

where

$$\hat{\lambda}_i = \lambda_i \cdot \frac{\pi}{180}, \qquad i \in \{1, 2\}$$
$$\hat{\phi}_i = \phi_i \cdot \frac{\pi}{180}, \qquad i \in \{1, 2\}$$

is the latitude and longitude of the points converted to radians.

Ideally, both to make the calculations easier and for better visualization and interpretation, we would prefer working on a plane rather than a sphere. For this purpose, we have converted the GPS coordinates to the Universal Transverse Mercator coordinate system (also known as UTM), hoping that we are working on a small enough area for the approximation to be adequate. We then calculated the distance between the two furthest points in the dataset using the UTM approximation and compared them to the actual distance calculated based on Lemma 5. The difference between the approximated distance and real distance was 0.13%, which is more than accurate enough for our purposes, especially considering this is the worst-case scenario and the distances which are the most interesting to us (i.e. the observations where the wolves are close together), are orders of magnitude smaller. Thus, from now on we will display all graphics/tables, as well as perform all calculations, in UTM coordinates rather than the original Latitude/Longitude.

We illustrate a simple chart of all the documented locations for each wolf in Figure 4.1. Based on the image, we have decided to eliminate two wolves from our analysis. The first one was Wolf 72, whose behavior seemed rather unusual compared to the other wolves, with its locations a lot more scattered well into the territories of Wolf 66 and Wolf 62. This can also be observed in the animation[2] on the project's website (for the sake of consistency we are using the same colors in our graphics that are used on the project website, with the exception of Wolf 72 for which we used black instead of white for obvious reasons). The other one was Wolf 62, whose behavior is a lot more standard, however we can clearly see the vast majority of its observations were far away from the boundary of its territory with only a few points near Wolf 71, so it does not seem productive to do much analysis on this pack.

[2]https://www.voyageurswolfproject.org/animations

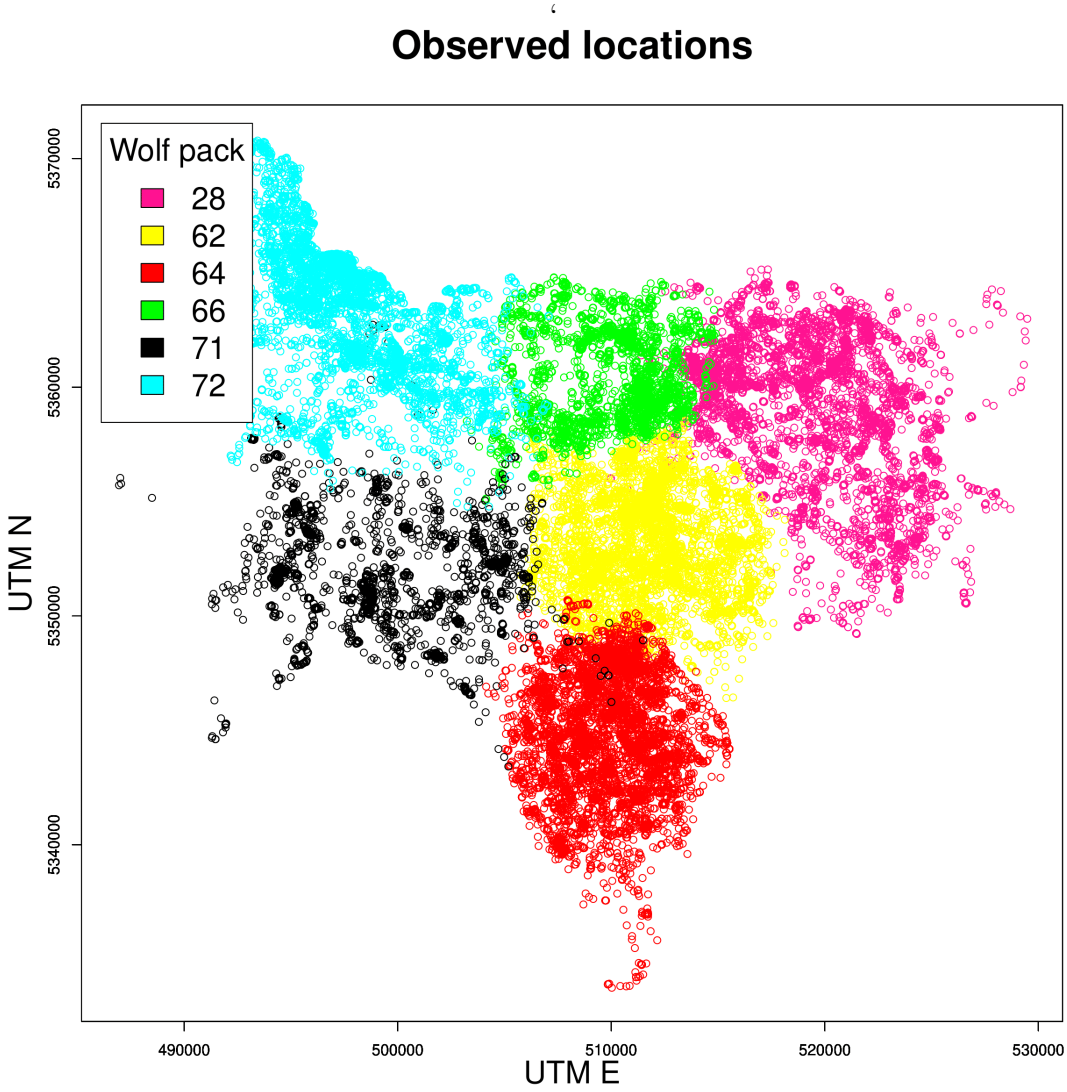


Figure 4.1: Map of all documented location of each wolf pack

As we mentioned before, the data was collected between April 15th 2018 and November 9th 2018. However, not all wolves were being tracked from the start to the end. In Figure 4.2 we show the interval during which each wolf was tracked, to get a better sense of the overlap. We can see that there are some discrepancies, especially with Wolf 71, however for the most part we have over 5 months worth of data for most wolf pairs.

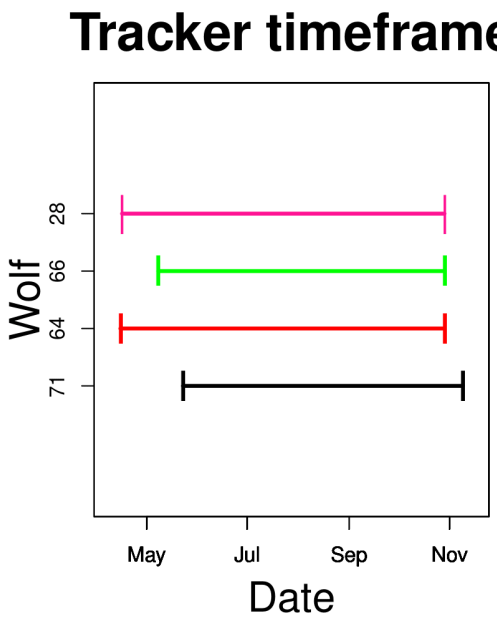


Figure 4.2: GPS tracker measured timeframe

Another issue we had to deal with was missing data within the timeframe itself. Some of the GPS tracker measurements were missing due to technical issues, so the measurements are not always exactly 20 minutes apart. This is especially an issue with Wolf 71 (and Wolf 72, but we have already excluded that one from our analysis). Of course, since we are interested in the distance between two wolves at a specific point in time, we can only use the measurements at times when *both* of the adjacent wolves' positions are known. From Figure 4.1 we can see there are only 4 adjacent pairs we could explore. In Table 4.1 we show the number of overlapping observations which are available to us for each of those 4 pairs.

There seem to be two main ways to deal with missing observations in our data. Let us start with a vector of the *incomplete* observations $((X_1, Y_1), ..., (X_n, Y_n))$. We need to do two separate things: We have to perform a transformation that restricts the data to a vector of observations at times where both X_i and Y_i are available, thus receiving a shorter vector $((\hat{X}_1, \hat{Y}_1), ..., (\hat{X}_m, \hat{Y}_m)), m \leq n$ (the important thing to note is that $\hat{X}_i$ was not necessarily measured at the same time as X_i, since we removed some observations). The second thing we have to do is, obviously, performing the "random" shift itself.

The two aforementioned solutions to missing data differ only in the order these two operations are applied in. The first option relies on creating the transformed vector first, and then applying the shift to it. The advantage of this approach is that each replication will have the same number of observations used to compute the test statistic, the disadvantage is that the observations aren't all shifted by the same time. The other solution, on the other hand, is to perform the shift first and then apply the transformation. The advantage here is that each observation is shifted by the same time, but if there are missing values in both processes (as is the case for our data), the sample size for each replication will be different.

Wolf #1	Wolf #2	Number of observations
28	64	12444
28	71	4596
64	66	10488
64	71	4521

Table 4.1: Number of observations for each adjacent wolf pair

Wolf #1	Wolf #2	p
28	64	0.994
28	71	0.932
64	66	0.077
64	71	0.024

Table 4.2: p-values using $S_{\text{mean},d=0.2}$ for each adjacent wolf pair

We have tested both options on the model from Chapter 2 in which we removed around 15% observations at random from both processes, and after comparing both options, we found the observed significance level to be practically the same for both. Thus we are choosing to apply the first option (i.e. finding the overlapping times first and then shifting one of the processes) and we are confident that this does not affect the Type I error.

4.2 Testing independence

Now we can move on to applying the tests developed in the previous chapter on the available data. We will be using the torus correction and the test statistics S_{mean} and S_{time}, since the previous chapter showed the functional test statistic does not seem to be worth using. The actual observation windows for the wolf packs appear to be very irregular, however this is not an issue for our torus correction since we are actually shifting the data in time rather than in space. It is important to note that with the sample sizes we have available, it is no longer feasible for us to perform every possible shift. We will therefore be limitting the number of shifts to 2000 (e.g. when we have 12000 observations we will perform 2000 shifts, shifting by 6 timestamps at a time).

We start with the test statistic S_{mean} with the parameter $d = 0.2$. The p-values for all 4 wolf pairs can be found in Table 4.2. The wolf pair $24, 71$ seems to be the most interesting, so we also provide an illustration of how the choice of d affects the p-value in Figure 4.3, where we can see that for higher values of d the p-value increases as we introduce less interesting observations into the test statistic compared to only utilizing the ones for which the wolves are close to each other. In Figure 4.4 we also show the histogram of the original measured distances and the distances in an example replication, we can see the value of the test statistic is higher for the replicated data, which indicates repelling interactions.

For S_{time}, the parameter selection is less obvious. In Chapter 3, we found that $c = \hat{I}$ was a good choice, however it is problematic in this case. Even if we were to

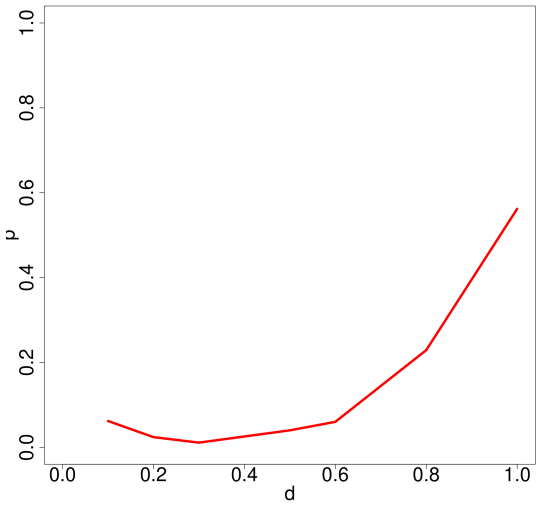


Figure 4.3: p-value of the test using S_{mean} on the wolf pair $64, 71$ based on the value of d

Wolf #1	Wolf #2	$\hat{I}$	Observations below $\hat{I}$
28	64	900	1
28	71	525	6
64	66	600	3
64	71	1725	1

Table 4.3: Estimates of I

assume the simplified behavioral model proposed in Chapter 3, the missing values are a bigger issues for the parameter estimation than they are for the Monte Carlo test. For instance, if there is a long time between observations, we are not really justified in treating these observations as subsequent. We will nevertheless attempt to compute an estimate $\hat{I}$ under these simplified assumptions. The results are shown in Table 4.3 along with the number of observations lower than this value. As we can see, the values $\hat{I}$ seem to be practically the smallest possible estimates for which we still have a nonzero number of observations. This provides some evidence for the independence of the wolf trajectories, as this behavior is very consistent with the simulation studies performed for independent processes in Chapter 3. The p-values for this test can be found in Table 4.4. Just like before, we provide a graph showing the dependence of the p-value for the most interesting wolf pair on the choice of the parameter c in Figure 4.5 and again we see that introducing more noise into the computation by increasing c increases the p-value significantly, which is exactly what we expected to see based on the results in Chapter 3.

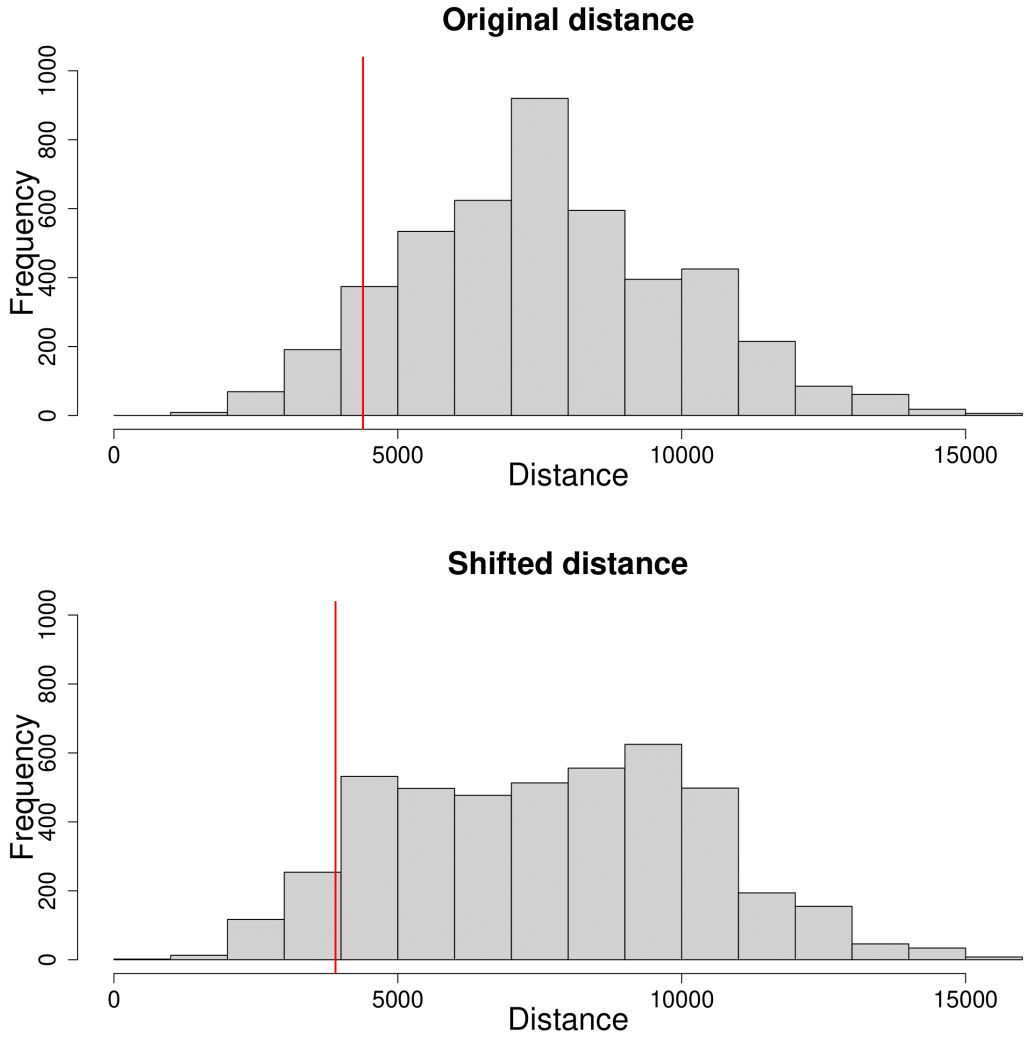


Figure 4.4: Distance distribution for the original data and for a selected shift ($K = 500$), red line shows the value of $S_{\text{mean},d=0.2}$

We can see the evidence for repelling behavior is far from convincing. Out of the eight performed tests (four wolf pairs, with two tests applied to each), only one showed a p-value $p < 0.05$. After applying any type of multiple testing correction, such as the Bonferroni correction (Bonferroni [1936]) or the Šidák correction (Šidák [1967]), we unfortunately cannot consider these results statistically significant. Thus our conclusion is that we do not reject the hypothesis of independence of any given wolf pair included in the analysis at a significance level 0.05.

Wolf #1	Wolf #2	p
28	64	0.802
28	71	0.887
64	66	0.676
64	71	0.061

Table 4.4: p-values using $S_{\text{time},c=\hat{I}}$ for each adjacent wolf pair

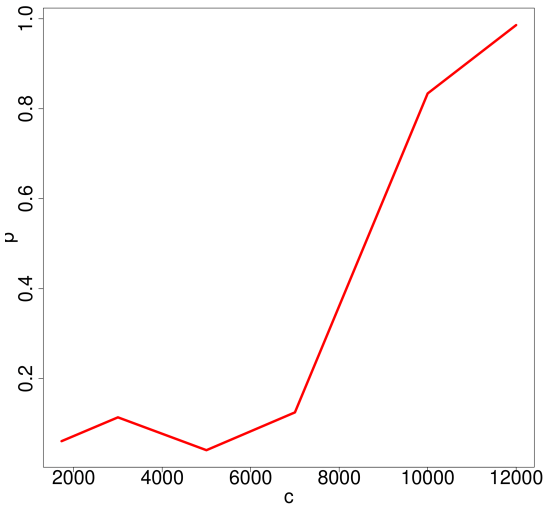


Figure 4.5: p-value of the test using S_{time} on the wolf pair $64, 71$ based on the value of c

Conclusion

In this thesis, we have dealt with nonparametric testing of the independence of animal trajectories. First, we proposed a parametric model suitable for simulating trajectories with repelling or attracting behavior. We discussed multiple generalizations of this model. Then, we derived the maximum likelihood estimates of the parameters of this model and performed a simulation study to assess the bias and mean squared error of these esimates based on the sample size and true values of the parameters.

Next, we covered the basics of Monte Carlo testing, as well as explained the theory behind envelope tests when using functional test statistics. We proposed three types of test statistics which could be used to test for dependence between two trajectories using the Monte Carlo test, with two of them being scalar and one of them being a functional test statistic based on the empirical distribution function. Then we performed a series of simulations using the previously designed model to assess the behavior of these statistics. We investigated how the achieved significance level and power of the tests changes with different sample size, dfferent model parameters and various generalizations of the model, as well as different tuning parameters of the test statistics.

Lastly, we have applied these tests to the data we had received from the Voyageurs Wolf Project to investigate whether we could find any repelling behavior between wolf packs in the Voyageurs National Park in Minnesota, USA. At the end of the testing, we have concluded that there is not enough evidence to reject the null hypothesis that the different wolf packs move independently of each other.

The tests developed in this thesis can further be extended in multiple ways. One of the main ones seems to be further research into the optimal choice of tuning parameters for the test statistics to ensure the highest possible power of the test. A more robust model could also be devised which mimics animal trajectories more accurately. Lastly, completely different choices of test statistics could always be considered. The statistics discussed in this thesis have been chosen in conjunction with the simplified trajectory model, and if there was an expectation of a vastly different type of interaction in the data, different test statistics can be better suited for the problem. We believe another area worth expanding upon are different functional test statistics.

Bibliography

J. Anděl. *Základy matematické statistiky*. Druhé opravené vydání. Matfyzpress, Praha, 2007. ISBN 80-7378-001-1.

Barnard. Discussion on professor bartlett's paper. *Journal of the Royal Statistical Society: Series B (Methodological)*, 25(2):281–296, 1963.

Julian Besag and Peter Clifford. Generalized monte carlo significance tests. *Biometrika*, 76(4):633–642, 1989. doi: 10.1093/biomet/76.4.633.

Julian Besag and Peter J. Diggle. Simple monte carlo tests for spatial pattern. *Applied Statistics*, 26(3):327, 1977. doi: 10.2307/2346974.

C. Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, 8:3–62, 1936. URL https://ci.nii.ac.jp/naid/20001561442/en/.

James Douglas Hamilton. *Time Series Analysis*. Princeton University Press, 1994. ISBN 978-0691042893.

S. Rao. Jammalamadaka and Ambar Sengupta. *Topics in circular statistics*. World Scientific, 2001.

N. Bert Loosmore and E. David Ford. Statistical inference using the g or k point pattern spatial statistics. *Ecology*, 87(8):1925–1931, 2006.

H. W. Lotwick and B. W. Silverman. Methods for analysing spatial processes of several types of points. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(3):406–413, 1982. ISSN 00359246.

Kanti V. Mardia. *Statistics of directional data*. Academic Press, 1972.

Kantilal V. Mardia and Peter E. Jupp. *Directional statistics*. Wiley, 2000. ISBN 0 471 95333 4.

Tomáš Mrkvička, Jiří Dvořák, Jonatan A. Gonzalez, and Jorge Mateu. Revisiting the random shift approach for testing in spatial statistics. 2019.

Mari Myllymäki, Tomáš Mrkvička, Pavel Grabarnik, Henri Seijo, and Ute Hahn. Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2):381–404, 2017. doi: https://doi.org/10.1111/rssb.12172. URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12172.

Nagy, Stanislav, Gijbels, Irène, Omelka, Marek, and Hlubinka, Daniel. Integrated depth for functional data: statistical properties and consistency. *ESAIM: PS*, 20:95–130, 2016. doi: 10.1051/ps/2016005.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL https://www.R-project.org/.

B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2):172–212, 1977. ISSN 00359246.

Rolf Schneider and Wolfgang Weil. *Stochastic and Integral Geometry*. Springer Berlin Heidelberg, 2008.

R. W. Sinnott. Virtues of the Haversine. *Sky and Telescope*, 68(2):158, December 1984.

Zbyněk Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967. ISSN 01621459. URL http://www.jstor.org/stable/2283989.